

LEMBAR PENGESAHAN

KLASIFIKASI RUMAH TANGGA MISKIN DI KABUPATEN JOMBANG BERDASARKAN FAKTOR-FAKTOR YANG MEMPENGARUHI DENGAN PENDEKATAN CART (*CLASSIFICATION* *AND REGRESSION TREES*)

TUGAS AKHIR

Diajukan Untuk Memenuhi Salah Satu Syarat Kelulusan
Program Studi S-1 Jurusan Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember

Oleh :

RIZA INAYAH
NRP 1312 105 018

Disetujui oleh Pembimbing Tugas Akhir

Dr. Bambang Widjanarko Otok, M.Si
NIP : 19681124 199412 1 001

Dr. Santi Wulan Purnami, M.Si
NIP : 19720923 199803 2 001

Mengetahui

Ketua Jurusan Statistika FMIPA-ITS

Dr. Muhammad Mashuri, MT
NIP. 19620408 198701 1 001

SURABAYA, JULI 2014



**KLASIFIKASI RUMAH TANGGA MISKIN DI KABUPATEN
JOMBANG BERDASARKAN FAKTOR-FAKTOR YANG
MEMPENGARUHI DENGAN PENDEKATAN CART
(CLASSIFICATION AND REGRESSION TREES)**

Nama Mahasiswa : Riza Inayah
NRP : 1312 105 018
Jurusan : Statistika FMIPA-ITS
Dosen Pembimbing : Dr. Bambang W. Otok, M.Si
Co. Pembimbing : Dr. Santi Wulan Purnami, M.Si,

ABSTRAK

Tingkat kemiskinan di Kabupaten Jombang dapat dikatakan relatif sulit bergerak turun dimana masih ada kurang lebih 73.720 Kepala Keluarga (KK) miskin dari 344 ribu KK yang tersebar di 302 desa dan 4 kelurahan. Pada penelitian ini ingin ditentukan klasifikasi status kemiskinan rumah tangga miskin di Kabupaten Jombang berdasarkan faktor-faktor yang diduga dominan mempengaruhi hasil klasifikasi dengan pendekatan CART (Classification and Regresion Trees). Unit penelitian adalah rumah tangga miskin di kabupaten Jombang sebanyak 73.720 rumah tangga. Variabel penelitian meliputi satu variable respon yaitu status kemiskinan rumah tangga miskin (Rumah Tangga Sangat Miskin atau RTSM, Rumah Tangga Miskin atau RTM, dan Rumah Tangga Hampir Miskin atau RTHM) dan ada delapan belas variabel prediktor. Penerapan pendekatan pohon klasifikasi menunjukkan bahwa variabel terpenting atau paling dominan berpengaruh dalam menentukan status kemiskinan suatu rumah tangga miskin pada penelitian ini yaitu besarnya penghasilan rata-rata per bulan (Rp) dengan skor tingkat kepentingan variabel sebesar 100. Keakuratan klasifikasi yang dihasilkan pohon optimal untuk data learning sebesar 40,986 persen sedangkan data testing sebesar 39,654 persen.

Kata Kunci: CART, Kabupaten Jombang, Klasifikasi Rumah Tangga Miskin

**CLASSIFICATION OF POOR HOUSEHOLDS IN JOMBANG
DISTRICT BASED ON INFLUENCES FACTOR USING
CART (CLASSIFICATION AND REGRESSION TREES)
APPROACH**

Name : Riza Inayah
NRP : 1312 105 018
Department : Statistika FMIPA-ITS
Counsellor : Dr. Bambang W. Otok, M.Si
Co. Counsellor : Dr. Santi Wulan Purnami, M.Si,

ABSTRACT

Poverty levels in Jombang District is relatively difficult to move down. There are about 73.720 Head of Households from 344 thousand families who are spread out in 302 villages and 4 sub-district. This research studied the poverty class levels of poor households in Jombang District based on dominantly significant factors which affect the classification using CART (Classification and Regression Trees) approach. Analyzing unit in this research is poor households as many as 73.720 households. The response variable of this research is three poverty class levels, they are very poor households (RTSM), poor households (RTM), and almost poor households (RTHM). There are eight teen predictor variables. The most important variable in optimal classification tree for classifying poor household levels in Jombang District is average income per month (Rp) with the score variable interest rate of it is 100. Classification accuracy produced by the tree optimal for learning data is 40.986 percent while 39.654 percent for testing data.

Key Words: CART, Jombang District, Poor Households Classification

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
PAGE TITLE	iii
LEMBAR PENGESAHAN	v
ABSTRAK	vii
ABSTRACT	ix
KATA PENGANTAR.....	xi
DAFTAR ISI	xiii
DAFTAR TABEL	xv
DAFTAR GAMBAR	xvii
DAFTAR LAMPIRAN	xix
BAB I PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan.....	4
1.4 Manfaat Penelitian	5
BAB II TINJAUAN PUSTAKA	
2.1 <i>Classification and Regression Trees (CART)</i>	7
2.1.1 Pembentukan Pohon Klasifikasi	11
2.1.2 Penentuan Ukuran Pohon Klasifikasi yang Layak	14
2.1.3 Pemangkasan Pohon Klasifikasi (<i>Pruning</i>)...	15
2.1.4 Penentuan Pohon Klasifikasi Optimum.....	17
2.1.5 Ilustrasi Pemilihan Pemilah Menggunakan Indeks Twoing	19
2.1.6 Ilustrasi Pemilihan Pemilah Menggunakan Indeks Gini	28
2.2 Teori Kemiskinan.....	42
2.3 Penelitian Sebelumnya.....	44
BAB III METODOLOGI PENELITIAN	
3.1 Sumber Data	47
3.2 Variabel Penelitian	47
3.3 Metode Analisis	54

BAB IV ANALISIS DAN PEMBAHASAN	
4.1 Karakteristik Rumah Tangga Miskin di Kabupaten Jombang Tahun 2010.....	59
4.2 Klasifikasi Rumah Tangga Miskin di Kabupaten Jombang dengan Pohon Klasifikasi CART	72
BAB V KESIMPULAN DAN SARAN	
5.1 Kesimpulan	87
5.2 Saran	88
DAFTAR PUSTAKA	89
LAMPIRAN	91
BIODATA PENULIS	

BAB II TINJAUAN PUSTAKA

2.1 *Classification and Regression Trees (CART)*

CART adalah singkatan dari *Classification and Regression Trees* yang mempunyai arti Pohon Klasifikasi dan Regresi. Maksudnya yaitu CART dapat menghasilkan suatu model pohon klasifikasi atau suatu pohon regresi tergantung dari variabel respon yang digunakan. CART termasuk salah satu metode statistik dengan teknik pohon keputusan untuk melakukan suatu analisis klasifikasi seperti halnya analisis regresi logistik dan analisis diskriminan yang merupakan contoh metode analisis klasifikasi klasik yang sering digunakan. Bedanya adalah analisis CART menggunakan pendekatan nonparametrik sedangkan analisis regresi logistik dan analisis diskriminan menggunakan pendekatan parametrik. CART akan menghasilkan suatu pohon klasifikasi apabila variabel respon berupa data kategorik, dan akan menghasilkan suatu pohon regresi apabila variabel respon berupa data kontinu.

Menurut Lewis (2000), masalah klasifikasi mencakup empat hal utama. Pertama adalah variabel respon kategorik. Variabel respon adalah karakteristik yang ingin diprediksi berdasarkan variabel prediktor. Kedua adalah variabel prediktor, yaitu karakteristik yang berpotensi memiliki hubungan atau mempengaruhi variabel respon. Pada umumnya ada beberapa variabel prediktor yang mungkin. Ketiga adalah data *learning*, yaitu suatu himpunan data yang mencakup nilai variabel respon dan variabel prediktor dari pengamatan untuk digunakan memprediksi hasil ke depan. Keempat adalah data *testing* yang terdiri atas pengamatan-pengamatan untuk digunakan mengetahui keakuratan prediksi hasil ke depan (mengevaluasi hasil prediksi). Dalam prakteknya, data testing bisa ada, bisa juga tidak. Namun secara umum diyakini bahwa data *testing* diperlukan untuk memvalidasi suatu keputusan klasifikasi. Masalah keputusan mencakup dua hal sebagai tambahan dari yang ditemukan pada masalah klasifikasi, antara lain bahwa untuk setiap pemilihan pengamatan secara random akan mempunyai probabilitas atau

kemungkinan suatu hasil tertentu. Selain itu, suatu pengamatan bisa jadi salah diklasifikasikan yang ditunjukkan dengan membuat suatu tabel perbandingan kondisi aktual dengan prediksi. Misal dari contoh Tabel 2.1 diperoleh informasi ada sebanyak n_{11} pengamatan kelas 1 yang tepat diklasifikasikan sebagai anggota kelas 1, ada sebanyak n_{12} pengamatan kelas 1 yang salah diklasifikasikan sebagai anggota kelas 2 dan ada sebanyak n_{13} pengamatan kelas 1 yang salah diklasifikasikan sebagai anggota kelas 3. Demikian seterusnya dengan interpretasi yang serupa untuk kelas 2 dan kelas 3.

Tabel 2.1 Contoh Tabel Prediksi Klasifikasi dengan 3 Level Kelas Variabel Respon

Kondisi Aktual	Prediksi Klasifikasi Pohon			Total
	Kelas 1	Kelas 2	Kelas 3	
Kelas 1	n_{11}	n_{12}	n_{13}	$n_{1.}$
Kelas 2	n_{21}	n_{22}	n_{23}	$n_{2.}$
Kelas 3	n_{31}	n_{32}	n_{33}	$n_{3.}$
Total	$n_{.1}$	$n_{.2}$	$n_{.3}$	N

Suatu analisis klasifikasi yang baik adalah yang mampu memberikan nilai keakuratan klasifikasi yang tinggi dengan tingkat kesalahan yang rendah. Perhitungan keakuratan klasifikasi perlu dilakukan untuk mengetahui seberapa besar pengelompokan sekumpulan data pengamatan dikategorikan dengan tepat pada suatu kelas atau kelompoknya. Salah satu ukuran evaluasi keakuratan hasil klasifikasi yaitu nilai 1-APER. APER (*Apparent Error Rate*) menunjukkan ukuran kesalahan klasifikasi sejumlah dataset pengamatan oleh suatu fungsi klasifikasi (Johnson dan Winchern, 1992). Mengacu pada Tabel 2.1, nilai keakuratan klasifikasi dengan menghitung nilai 1-APER dapat dituliskan dalam persamaan rumus berikut.

$$AkurasiKlasifikasi = 1 - APER = \frac{[n_{11} + n_{22} + n_{33}]}{N} \quad (2.1)$$

CART mempunyai beberapa kelebihan, antara lain yaitu hasilnya lebih mudah diinterpretasikan karena struktur pohon klasifikasi yang dihasilkan dapat dilihat secara visual, proses pengklasifikasian lebih mudah dilakukan dengan menelusuri

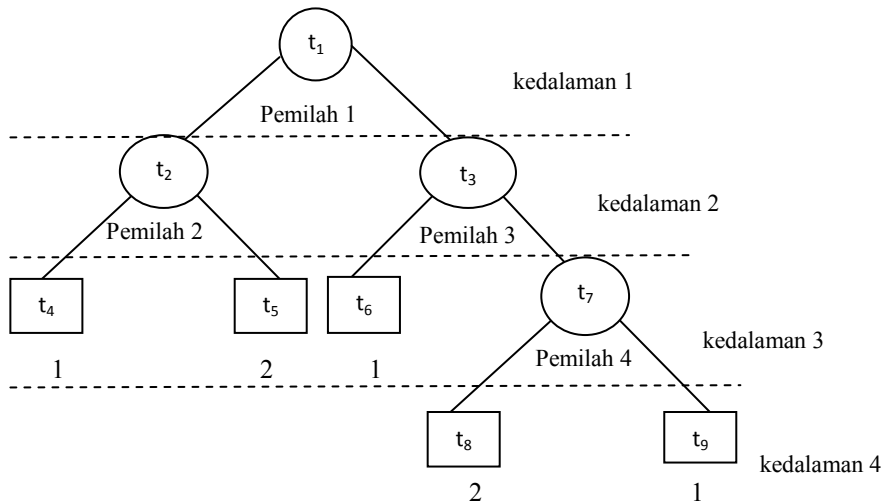
pohon klasifikasi yang dihasilkan. Selain itu, CART menggunakan pendekatan nonparametrik yang tidak memerlukan asumsi tertentu sehingga memberikan fleksibilitas yang lebih besar dalam menganalisa berbagai jenis data yang ingin diterapkan, termasuk himpunan data yang mempunyai jumlah besar, mengandung variabel yang sangat banyak dan mencakup skala data campuran (kategorik dan kontinu). Disamping kelebihan, metode CART juga mempunyai kekurangan jika dibandingkan dengan analisis klasifikasi klasik seperti regresi logistik dan analisis diskriminan, yaitu kurang stabil pada perubahan data *learning* yang digunakan, dimana itu bisa menyebabkan perubahan besar pada hasil prediksi pohon klasifikasi (Timofeev, 2004 diacu dalam Muttaqin, 2013).

Struktur CART diperoleh melalui penerapan suatu prosedur penyekatan biner yang dapat dilakukan secara berulang kali untuk membentuk partisi-partisi kelas pengamatan yang lebih homogen dengan karakteristik tertentu. Prosedur tersebut dikenal dengan sebutan *binary recursive partitioning*. Istilah “*binary*” mengimplementasikan bahwa masing-masing kelompok pengamatan ditampilkan oleh suatu *node* atau simpul dalam suatu pohon keputusan yang hanya bisa dipilah menjadi dua kelompok. Struktur CART diawali dengan pembentukan simpul utama (*root node*) sebagai pemilah terpenting dalam menentukan klasifikasi kelas pengamatan, yang menghimpun seluruh pengamatan dari data *learning* yang digunakan. Simpul utama tersebut sekaligus menjadi simpul induk (*parent node*) pertama yang hanya akan bisa dipilah menjadi dua simpul baru atau disebut simpul anak (*child node*) kiri dan kanan. Jadi, setiap simpul hanya bisa dipilah menjadi dua simpul baru atau simpul anak, yang mana simpul awal tersebut disebut sebagai simpul induk dari simpul anak kiri dan kanan yang baru. Istilah “*recursive*” mengacu pada kenyataan bahwa proses pemilahan secara biner tersebut bisa diulang-ulang seterusnya yaitu memilah simpul induk menjadi dua simpul anak kiri dan kanan kemudian masing-masing simpul anak tersebut akan berperan sebagai simpul induk baru yang akan dipilah lagi menjadi dua simpul anak baru (selama masih memungkinkan untuk dipilah). Seperti demikian seterusnya bisa diulang-ulang sampai terbentuk simpul terminal yaitu simpul akhir yang sudah

tidak bisa dipilah lagi. Istilah ”*partitioning*” menunjukkan bahwa himpunan data awal dipilah menjadi bagian-bagian atau partisi (Lewis, 2000).

Langkah-langkah algoritma CART secara umum meliputi hal-hal berikut.

- Pembentukan pohon klasifikasi
- Penghentian proses pembentukan pohon klasifikasi
- Pemangkasan pohon klasifikasi
- Penentuan pohon klasifikasi optimal



Gambar 2.1 Contoh struktur pohon klasifikasi dengan dua level kelas variabel respon.

Pada contoh struktur pohon klasifikasi yang ditampilkan Gambar 2.1, simpul utama dinotasikan dengan t_1 . Variabel prediktor yang menjadi pemilah simpul utama merupakan variabel terpenting yang mampu memisahkan kelas (kategori / level variabel respon) dengan jumlah pengamatan terbanyak. Sementara t_2 , t_3 , dan t_7 merupakan simpul dalam (*internal nodes*). Sedangkan t_4 , t_5 , t_6 , t_8 dan t_9 merupakan simpul terminal (*terminal node*) karena pada simpul tersebut tidak terjadi pemilahan simpul anak lagi, sebagaimana pada Gambar 2.1 dilambangkan dengan bentuk segi empat. Perhitungan kedalaman (*depth*) dimulai dari

simpul utama (t_1) yang berada pada kedalaman 1, sedangkan t_2 dan t_3 pada kedalaman 2, begitu seterusnya sampai pada simpul terminal t_8 dan t_9 yang berada dalam kedalaman 4. Sehingga akan diperoleh sekatan-sekatan dengan respon yang lebih homogen dalam tiap sekatan. Simpul terminal nantinya masing-masing akan ditandai dengan label kelas. Simpul-simpul terminal dengan label kelas yang sama akan diklasifikasikan ke dalam suatu himpunan bagian tertentu sehingga dapat diketahui karakteristik dari setiap klasifikasi kelas yang terbentuk, misalnya untuk contoh di atas akan diperoleh karakteristik pengamatan yang terklasifikasi kelas 1 berdasarkan $A_1 = t_4 \cup t_6 \cup t_9$ dan karakteristik pengamatan yang terklasifikasi kelas 2 berdasarkan $A_2 = t_5 \cup t_8$.

Berikut diuraikan tahapan yang dilakukan dalam analisis pohon klasifikasi CART dengan variabel respon kategorik.

2.1.1 Pembentukan Pohon Klasifikasi

Proses pembentukan pohon klasifikasi perlu memperhatikan 3 hal penting antara lain memilih pemilah, memutuskan apakah simpul perlu dipilah lagi atau sudah dijadikan simpul terminal, dan memberi label kelas pada setiap simpul terminal yang terbentuk (Breiman *et al.*, 1993).

Proses pembentukan pohon klasifikasi membutuhkan data *learning* sehingga sebelumnya perlu dicari dulu metode terbaik untuk pembentukan pohon klasifikasi yaitu yang menghasilkan ketepatan klasifikasi data *testing* tertinggi. Dengan demikian berarti data keseluruhan perlu dibagi menjadi dua himpunan terlebih dulu menjadi L_1 (data *learning*) dan L_2 (data *testing*). Data *learning* yang masih bersifat heterogen berada dalam suatu ruang atau simpul. Simpul tersebut menjadi simpul utama yang perlu dipilah oleh salah satu variabel prediktor sebagai pemilah utama. Pemilihan pemilah dimulai dengan memeriksa nilai-nilai variabel prediktor pada tiap simpul dengan dua tahap yaitu dengan mencari semua kemungkinan pemilah pada setiap variabel prediktor dan mencari pemilah terbaik dari setiap variabel prediktor. Aturan pemilihan menurut Breiman *et al.* (1993) yaitu setiap pemilah hanya bergantung pada nilai yang berasal dari satu

variabel prediktor. Berikut ini beberapa ketentuan nilai pemilah pada masing-masing jenis skala variabel.

- Untuk variabel kontinu, jika ruang sampel berukuran n dan terdapat n nilai amatan yang berbeda pada variabel X_j , maka akan terdapat sebanyak $n-1$ pemilah yang berbeda.
- Untuk variabel kategorik, jika variabel X_j adalah kategori nominal bertaraf L , maka akan ada sebanyak $2^{L-1}-1$ kemungkinan pemilah.
- Untuk variabel kategorik, jika variabel X_j adalah kategori ordinal maka akan ada sebanyak $L-1$ pemilah yang mungkin.

Tujuan dilakukan pemilahan adalah untuk mengurangi keheterogenan pada simpul utama dan mendapatkan simpul anak-simpul anak dengan tingkat homogenitas yang tinggi dan dapat mencirikan karakteristik kelas-kelas variabel respon. Untuk mengetahui kondisi tersebut, maka ada implementasi fungsi keheterogenan simpul. Fungsi keheterogenan yang bisa diterapkan menurut Breiman *et al.* (1993) adalah indeks Gini dan indeks Twoing.

- Fungsi keheterogenan indeks Gini sebagai berikut.

$$i(t) = \sum_{i \neq j} p(i | t) p(j | t) \quad (2.2)$$

dengan

$i(t)$ = Fungsi keheterogenan simpul t

$p(i|t)$ = Proporsi kelas i pada simpul t

$p(j|t)$ = Proporsi kelas j pada simpul t

- Fungsi kehetoreganan indeks Twoing sebagai berikut.

$$i(t) = \frac{p_L p_R}{4} \left[\sum_j |p(j | t_L) - p(j | t_R)| \right]^2 \quad (2.3)$$

dengan

$i(t)$ = Fungsi keheterogenan simpul t

p_L = Proporsi pengamatan pada simpul kiri

p_R = Proporsi pengamatan pada simpul kanan

$p(j|t_L)$ = Proporsi pengamatan dari simpul t menuju simpul kiri dengan kelas j

$p(j|t_R)$ = Proporsi pengamatan dari simpul t menuju simpul kanan dengan kelas j

Kemudian pemilihan pemilah terbaik dilakukan berdasarkan kriteria *goodness of split* yang merupakan suatu evaluasi pemilahan oleh suatu pemilah s pada suatu simpul t . *Goodness of split* menunjukkan ukuran penurunan keheterogenan suatu kelas atau kualitas ukuran seberapa baik pemilah s menyaring data menurut kelas agar lebih homogen. Misal terdapat pemilah s yang akan memilah t menjadi simpul kiri (t_L) dengan proporsi p_L dan simpul kanan (t_R) dengan proporsi p_R , maka *goodnes of split* yang secara matematis dinotasikan dengan $\phi(s, t)$ didefinisikan sebagai fungsi penurunan keheterogenan sebagai berikut.

$$\phi(s, t) = \Delta i(s, t) = i(t) - p_L i(t_L) - p_R i(t_R) \quad (2.4)$$

dengan

$\phi(s, t)$ = Nilai *Goodness of split*

$i(t)$ = Fungsi keheterogenan simpul t

$p_L i(t_L)$ = Proporsi pengamatan dari simpul t menuju simpul kiri

$p_R i(t_R)$ = Proporsi pengamatan dari simpul t menuju simpul kanan

Pembentukan struktur pohon dilakukan dengan pencarian semua pemilah yang mungkin pada setiap simpul yang dimulai dari simpul utama t_1 . Suatu pemilah s^* yang mampu menghasilkan nilai penurunan keheterogenan tertinggi diantara semua pemilah s yang mungkin adalah yang terpilih sebagai pemilah simpul utama t_1 menjadi simpul anak t_2 dan t_3 . Konsep tersebut secara matematis dapat dituliskan dalam bentuk persamaan $\Delta i(s^*, t_1) = \max_{s \in S} \Delta i(s, t_1)$.

Kemudian untuk pengembangan pembentukan struktur pohon maka dengan cara dan konsep yang sama dilakukan pemilahan pada simpul t_2 dan t_3 masing-masing secara terpisah, seterusnya sampai ditemukan simpul terminal.

Suatu simpul t akan diputuskan untuk dipilah lagi ataukah menjadi simpul terminal atau dengan kata lain tidak akan dipilah lagi adalah tergantung daripada kondisi simpul tersebut. Apabila

kondisi simpul tersebut memenuhi salah satu kriteria berikut maka simpul tersebut akan menjadi simpul terminal.

- a) Hanya ada satu pengamatan ($n = 1$) dalam tiap simpul anak atau adanya batasan minimum n pengamatan yang diinginkan peneliti.
- b) Semua pengamatan dalam setiap simpul anak mempunyai distribusi yang identik terhadap variabel prediktor sehingga tidak mungkin untuk dipilih lagi.
- c) Adanya batasan jumlah level atau tingkat kedalaman pohon maksimal yang ditetapkan peneliti.

Apabila struktur pohon telah terbentuk mulai dari simpul utama t_i sampai dengan simpul terminal dimana sudah tidak lagi ditemukan simpul yang perlu dipilah lagi maka itu berarti pohon klasifikasi maksimal telah terbentuk. Pohon klasifikasi maksimal yaitu pohon klasifikasi yang memiliki jumlah simpul paling banyak. Setiap simpul terminal perlu diberi label kelas sehingga nantinya dapat diketahui karakteristik dari klasifikasi pengamatan untuk setiap kelas variabel respon yang terbentuk. Pemberian label kelas pada simpul terminal dilakukan berdasarkan aturan jumlah anggota kelas terbanyak, yaitu apabila,

$$p(j_0 | t) = \max_j p(j | t) = \max_j \frac{N_j(t)}{N(t)} \quad (2.5)$$

dengan

$p(j|t)$ = Proporsi kelas j pada simpul t

$N_j(t)$ = Jumlah pengamatan kelas j pada simpul t

$N(t)$ = Jumlah seluruh pengamatan pada simpul t

maka label kelas untuk simpul terminal t adalah j_0 . Dengan mengasumsikan biaya kesalahan klasifikasi untuk setiap kelas besarnya sama, pemberian label kelas j_0 tersebut memberikan nilai dugaan kesalahan pengklasifikasian pada simpul t paling kecil yaitu sebesar $r(t) = 1 - \max_j p(j | t)$.

2.1.2 Penentuan Ukuran Pohon Klasifikasi yang Layak

Pada tahap pertama sudah terbentuk suatu pohon klasifikasi maksimal. Pada umumnya, pohon klasifikasi maksimal yang telah terbentuk dengan simpul terminal paling banyak bisa jadi akan mengakibatkan terjadinya kasus *underfit* (nilai yang dihasilkan di

bawah kenyataan) ataupun *overfit* (nilai yang dihasilkan melebihi kenyataan yang ada) jika ukuran pohon tersebut belum *fit* atau layak. Ini kemungkinan karena adanya pengaruh dari aturan penghentian pemilahan simpul terminal (*stopping tree building rules*) yang hanya didasarkan pada banyaknya pengamatan pada simpul terminal ataupun besarnya penurunan keheterogenan. Semakin banyak pemilahan yang dilakukan dapat mengakibatkan semakin kecil tingkat kesalahan prediksi, sebab hal itu memungkinkan simpul terminal beranggotakan hanya satu pengamatan. Sementara apabila pengembangan pohon dibatasi dengan ketepatan batas tertentu, kasus *underfit* bisa saja terjadi. Untuk menghindari terbentuknya pohon klasifikasi yang terlalu besar dan kompleks maka perlu dilakukan proses pemangkasan pohon (*pruning*) agar pohon yang dihasilkan lebih layak.

2.1.3 Pemangkasan Pohon Klasifikasi (*Pruning*)

Pemangkasan pohon klasifikasi yaitu suatu penilaian ukuran pohon tanpa mengorbankan ketepatan atau kebaikannya melalui pengurangan simpul pohon yang dianggap tidak begitu signifikan berarti sehingga dicapai ukuran pohon yang layak. Secara berurutan atau bertahap untuk menghasilkan sebuah rangkaian pohon yang sederhana dan lebih sederhana, digunakan metode pemangkasan *cost complexity*. Metode ini mengandalkan *parameter complexity* atau ukuran kompleksitas yang dinotasikan α , yang secara perlahan nilainya meningkat selama proses pemangkasan. Dimulai pada level terakhir (simpul terminal), suatu simpul anak akan dipangkas jika hasil perubahan nilai kesalahan klasifikasi prediksi lebih kecil daripada α kali perubahan kompleksitas pohon. Dengan demikian, α adalah ukuran dari berapa banyak tambahan akurasi suatu pemilah harus menambah seluruh pohon untuk menjamin kompleksitas tambahan. Seiring peningkatan α , semakin banyak simpul dipangkas jauh, sehingga pohon menjadi lebih sederhana dan lebih sederhana.

Pada sembarang pohon T yang merupakan sub pohon dari pohon klasifikasi maksimal T_{max} ($T < T_{max}$), dengan nilai $\alpha \geq 0$, maka persamaan fungsi *cost complexity* yaitu.

$$R_a(T) = R(T) + \alpha |\tilde{T}| \quad (2.6)$$

dengan

$R_a(T)$ = *Cost complexity measure* (ukuran kompleksitas) suatu pohon T pada kompleksitas α

$R(T)$ = *Cost misclassification T tree* (ukuran kesalahan klasifikasi pohon T)

α = *Complexity parameter* (kompleksitas parameter) oleh penambahan satu simpul akhir pada pohon T

$|\tilde{T}|$ = Banyaknya simpul terminal pohon T

$R_a(T)$ merupakan kombinasi linear dari harga dan nilai kompleksitas pohon yang dibentuk. *Cost complexity pruning* menentukan suatu pohon bagian $T(\alpha)$ yang meminimumkan $R_a(T)$ pada seluruh pohon bagian untuk setiap nilai α . Selanjutnya pencarian pohon bagian $T(\alpha) < T_{max}$ yang meminimumkan $R_a(T)$ yaitu $R_a(T(\alpha)) = \min_{T < T_{max}} R_a(T)$.

Prosedur pemangkasan pohon klasifikasi secara umum dapat dituliskan sebagai berikut.

- i. Pohon klasifikasi maksimal T_{max} terbentuk lalu ambil simpul anak kiri t_L dan simpul anak kanan t_R yang merupakan hasil pemilahan simpul induk t .
- ii. Jika dari langkah ke-i. di atas dipenuhi persamaan $R(t) = R(t_L) + R(t_R)$ maka simpul anak t_L dan t_R dipangkas. Hasilnya yaitu pohon T_I yang memenuhi $R(T_I) = R(T_{maks})$
- iii. Lakukan kembali seperti langkah i dan ii sampai tidak ada lagi pemangkasan yang terjadi. Hasil pemangkasan tersebut akan berupa suatu rangkaian urutan atau barisan menurun pohon yaitu $T_1 > T_2 > \dots > \{t_1\}$ dengan $T_1 < T_{max}$ dan secara bersesuaian ada rangkaian yang naik dari parameter *cost complexity* yaitu $\alpha_1 = 0 < \alpha_2 < \dots$ sebagaimana untuk $\alpha_k \leq \alpha < \alpha_{k+1}$, T_k adalah sub-pohon terkecil dari T_{max} yang meminimumkan $R_a(T)$

2.1.4 Penentuan Pohon Klasifikasi Optimal

Ada dua jenis penduga pengganti untuk mendapatkan pohon klasifikasi optimal yaitu:

a. *Test Sample Estimate*

Pada *test sample estimate*, sampel keseluruhan total L data dibagi menjadi dua himpunan yaitu L_1 (data *learning*) dan L_2 (data *testing*). Pembagian data *learning* dan data *testing* bisa diproporsikan sesuai peneliti namun dianjurkan pada data *learning* jumlahnya lebih banyak dari pada data *testing* karena pengamatan dalam L_1 digunakan untuk membentuk model pohon T , sedangkan pengamatan pada L_2 digunakan untuk menduga $R(T)$. Jika $N^{(2)}$ adalah jumlah pengamatan dalam L_2 , maka *test sample estimate* dihitung sebagai berikut.

$$R^{ts}(T_k) = \frac{1}{N^{(2)}} \sum_{i,j} C(i|j) N_{ij}^{(2)} \quad (2.7)$$

dengan

$R^{ts}(T_k)$ = Total proporsi kesalahan *test sample estimate*

$N^{(2)}$ = Jumlah pengamatan dari L_2

$C(i|j)$ = Jumlah Proporsi ke- i dan ke- j dari L_2

$N_{ij}^{(2)}$ = Jumlah kelas ke- j dalam L_2 dimana klasifikasi prediksinya adalah kelas ke- i

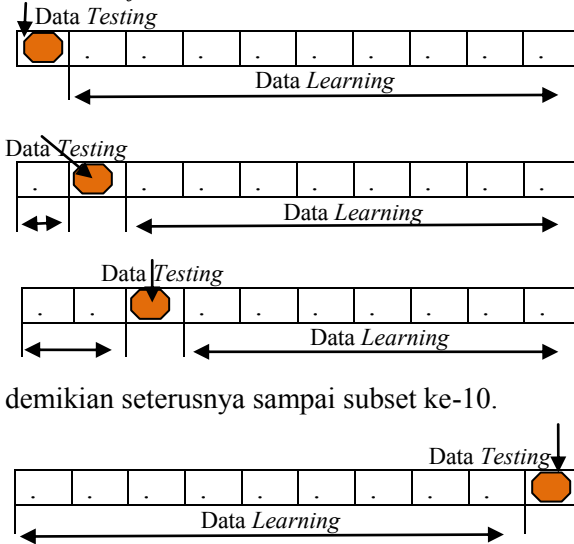
Pohon klasifikasi optimal yang dipilih yaitu T_k dengan

$$R^{ts}(T_k) = \min_t R^{ts}(T_k)$$

b. *Cross Validation V-fold Estimate*

Cross validation membagi data secara acak menjadi V subset yang berukuran relatif sama. Salah satu subset dicadangkan sebagai data *testing* dan subset-subset sisanya digabung dijadikan sebagai data *learning* dalam prosedur pembentukan model. Seluruh prosedur pembentukan model diulang V kali, dengan subset berbeda dari data setiap kali melakukan pembentukan pohon (Lewis, 2000). Metode *cross validation* bisa dicoba untuk menghindari *overlapping* pada data *testing*. Nilai V yang seringkali dipakai dan dijadikan standar adalah 10 atau berarti *cross validation 10-fold*. Sebab hasil dari berbagai percobaan yang ekstensif dan

pembuktian teoritis, menunjukkan bahwa *cross validation 10-fold* adalah pilihan terbaik untuk mendapatkan hasil validasi yang akurat. Berikut ini ilustrasi prosedur *cross validation 10-fold*.



demikian seterusnya sampai subset ke-10.

Gambar 2.2 Contoh ilustrasi prosedur *cross validation 10-fold*.

Cross validation V-fold estimate untuk T_k yang menggunakan pengamatan L dalam membentuk deretan pohon $\{T_k\}$ adalah sebagai berikut.

$$R^{cv}(T_k(\alpha)) = \frac{1}{N} \sum_{i,j} C(i|j) N_{ij} \quad (2.8)$$

dengan

$R^{cv}(T_k(\alpha))$ = Total proporsi T Cross validation v -fold estimate

$C(i|j)$ = Jumlah proporsi ke- i dan ke- j dari keseluruhan data pengamatan

N_{ij} = Jumlah kelas ke- i dan ke- j dari keseluruhan data pengamatan

Pohon klasifikasi optimal yang dipilih yaitu T_k dengan

$$R^{cv}(T_k) = \min_k R^{cv}(T_k)$$

2.1.5 Ilustrasi Pemilihan Pemilah Menggunakan Indeks Twoing

Berikut ini diberikan sebuah ilustrasi pembentukan pohon klasifikasi dengan fungsi keheterogenan menggunakan aturan indeks Twoing. Variabel yang digunakan meliputi satu variabel respon dan dua variabel prediktor. Masing-masing kriteria variabel tersebut secara lebih rinci ditampilkan dalam Tabel 2.2 sebagai berikut.

Tabel 2.2 Variabel untuk Ilustrasi I

	Variabel	Skala Data	Kategori atau Kelas
Y	Status Rumah Tangga (RT)	Ordinal	1. RT sangat miskin 2. RT miskin 3. RT hampir miskin
X ₁	Jenis dinding terluas	Nominal	1. Tembok 2. Kayu
X ₂	Luas lantai (m ²)	Rasio	-

Ada 10 data pengamatan diberikan sebagai data ilustrasi yang akan diolah sesuai dengan algoritma CART. Data tersebut ditampilkan dalam Tabel 2.3.

Tabel 2.3 Data Ilustrasi I

No. Sampel Pengamatan	Y	X₁	X₂
1	3	1	50
2	1	1	40
3	1	2	30
4	2	2	40
5	3	1	60
6	2	2	50
7	1	1	40
8	3	1	50
9	2	2	30
10	3	1	50

Pada awalnya, seluruh data sampel pengamatan tersebut berada dalam satu himpunan yang masih bersifat heterogen.

Algoritma CART yang pertama yaitu pembentukan pohon klasifikasi, diawali dengan pemilihan pemilah untuk simpul utama. Pemilihan pemilah dimulai dengan menghitung indeks Twoing sesuai persamaan (2.3).

Variabel X_1 terdiri atas 2 level kategori berskala nominal sehingga kemungkinan pemilah ada sebanyak $2^{2-1} - 1 \approx 1$ kemungkinan yaitu pemilah $X_1=1$ sebagai simpul kiri dan $X_1=2$ sebagai simpul kanan.

Tabel 2.4 Proporsi Pengamatan ke Simpul Kiri dan Kanan Jika Dipilah oleh X_1

		$N_L = 6$	$p_L = \frac{6}{10}$
Simpul kiri	Kelas 1	$N(1 t_L) = 2$	$p(1 t_L) = \frac{2}{6}$
	Kelas 2	$N(2 t_L) = 0$	$p(2 t_L) = 0$
	Kelas 3	$N(3 t_L) = 4$	$p(3 t_L) = \frac{4}{6}$
		$N_R = 4$	$p_R = \frac{4}{10}$
Simpul kanan	Kelas 1	$N(1 t_R) = 1$	$p(1 t_R) = \frac{1}{4}$
	Kelas 2	$N(2 t_R) = 3$	$p(2 t_R) = \frac{3}{4}$
	Kelas 3	$N(3 t_R) = 0$	$p(3 t_R) = 0$

Ket: L = *Left* (kiri) ; R = *Right* (kanan)

$$\begin{aligned}
 i(t) &= \frac{p_L p_R}{4} \left[\sum_j |p(j | t_L) - p(j | t_R)| \right]^2 \\
 &= \frac{(0.6)(0.4)}{4} \left[\left| \left(\frac{2}{6} - \frac{1}{4} \right) \right| + \left| \left(0 - \frac{3}{4} \right) \right| + \left| (0 - 0) \right| \right]^2 \\
 &= 0.06 [0.083 + 0.75 + 0]^2 \\
 &= 0.04163
 \end{aligned}$$

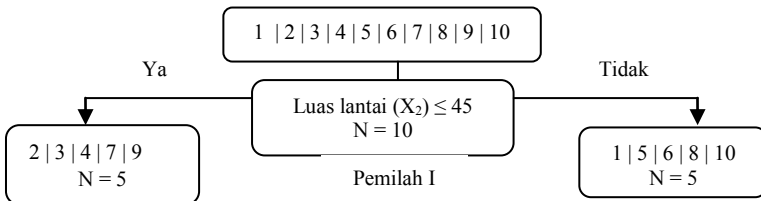
Variabel X_2 berskala rasio. Maka perlu ditentukan *threshold* pemilah dengan menghitung nilai median data. Diperoleh median data sebesar 45. Maka, data X_2 yang ≤ 45 akan dipilah ke simpul kiri sedangkan data X_2 yang > 45 dipilah ke simpul kanan.

Tabel 2.5 Proporsi Pengamatan ke Simpul Kiri dan Kanan Jika Dipilah *Threshold* $X_2 \leq 45$

		$N_L = 5$	$p_L = \frac{5}{10}$
Simpul kiri	Kelas 1	$N(1 t_L) = 3$	$p(1 t_L) = \frac{3}{5}$
	Kelas 2	$N(2 t_L) = 2$	$p(2 t_L) = \frac{2}{5}$
	Kelas 3	$N(3 t_L) = 0$	$p(3 t_L) = 0$
		$N_R = 5$	$p_R = \frac{5}{10}$
Simpul kanan	Kelas 1	$N(1 t_R) = 0$	$p(1 t_R) = 0$
	Kelas 2	$N(2 t_R) = 1$	$p(2 t_R) = \frac{1}{5}$
	Kelas 3	$N(3 t_R) = 4$	$p(3 t_R) = \frac{4}{5}$

$$\begin{aligned}
 i(t) &= \frac{p_L p_R}{4} \left[\sum_j |p(j | t_L) - p(j | t_R)| \right]^2 \\
 &= \frac{(0.5)(0.5)}{4} \left[\left| \left(\frac{3}{5} - 0 \right) \right| + \left| \left(\frac{2}{5} - \frac{1}{5} \right) \right| + \left| \left(0 - \frac{4}{5} \right) \right| \right]^2 \\
 &= 0.0625 [0.6 + 0.2 + 0.8]^2 = 0.16
 \end{aligned}$$

Aturan Twoing memilih pemilah yang memaksimumkan nilai $i(t)$. Dari perhitungan, nilai $i(t)$ yang dihasilkan pemilah $X_2 \leq 45$ lebih besar daripada nilai $i(t)$ yang dihasilkan pemilah $X_1 = 1$. Maka sebagai pemilah simpul utama dipilih $X_2 \leq 45$.



Gambar 2.3 Contoh pohon klasifikasi ilustrasi I hasil pemilahan I.

Simpul kiri dan simpul kanan yang dihasilkan dari proses pemilahan I masih bersifat heterogen sehingga perlu dilakukan pemilahan kembali. Berikut adalah data anggota simpul kiri.

Tabel 2.6 Data Anggota Simpul Kiri Hasil Pemilah I

No. Sampel Pengamatan	Y	X ₁	X ₂
2	1	1	40
3	1	2	30
4	2	2	40
7	1	1	40
9	2	2	30

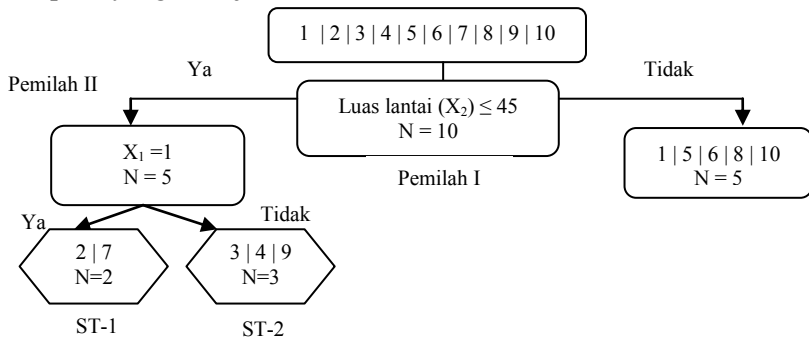
Simpul kiri hasil pemilah I akan dipilah menjadi simpul anak kiri dan kanan yang baru oleh pemilah II. Berikut ini perhitungan kemungkinan pemilah II.

Tabel 2.7 Proporsi Pengamatan dari Simpul Kiri Hasil Pemilah I ke Simpul Kiri dan Kanan Jika Dipilah oleh X₁

		$N_L = 2$	$p_L = \frac{2}{5}$
Simpul kiri X ₁ =1	Kelas 1	$N(1 t_L) = 2$	$p(1 t_L) = \frac{2}{2} = 1$
	Kelas 2	$N(2 t_L) = 0$	$p(2 t_L) = 0$
	Kelas 3	$N(3 t_L) = 0$	$p(3 t_L) = 0$
		$N_R = 3$	$p_R = \frac{3}{5}$
Simpul kanan X ₁ =2	Kelas 1	$N(1 t_R) = 1$	$p(1 t_R) = \frac{1}{3}$
	Kelas 2	$N(2 t_R) = 2$	$p(2 t_R) = \frac{2}{3}$
	Kelas 3	$N(3 t_R) = 0$	$p(3 t_R) = 0$

$$\begin{aligned}
 i(t) &= \frac{p_L p_R}{4} \left[\sum_j |p(j | t_L) - p(j | t_R)| \right]^2 \\
 &= \frac{(0.4)(0.6)}{4} \left[\left| \left(1 - \frac{1}{3}\right) - \left(0 - \frac{2}{3}\right) \right| + |(0 - 0)| \right]^2 \\
 &= 0.06 [0.67 + 0.67 + 0]^2 = 0.1067
 \end{aligned}$$

Dari Tabel 2.6 dapat dihitung median dari X_2 sebagai *threshold* pemilah X_2 , diperoleh senilai 40. Maka, data X_2 yang ≤ 40 akan dipilah ke simpul kiri sedangkan yang > 40 dipilah ke simpul kanan. Namun ternyata semua pengamatan hanya akan masuk ke simpul kiri. Oleh sebab itu, pemilah $X_2 \leq 40$ kurang tepat jika digunakan sebagai pemilah karena nanti hasilnya akan sama saja dengan simpul sebelumnya. Dengan demikian simpul kiri akan dipilah oleh pemilah dari $X_1=1$. Simpul kiri hasil pemilah II beranggotakan 2 pengamatan sementara simpul kanan beranggotakan 3 pengamatan. Kedua simpul anak baru tersebut dianggap homogen sehingga sudah tidak perlu dipilah kembali. Simpul kiri baru tersebut kemudian dijadikan sebagai simpul terminal ke-1 (ST-1) dan simpul kanannya sebagai simpul terminal ke-2 (ST-2). Pohon klasifikasi yang terbentuk menjadi seperti yang ditunjukkan oleh Gambar 2.4.



Gambar 2.4 Contoh pohon klasifikasi ilustrasi I sampai hasil pemilahan II.

Sebagaimana simpul kiri, simpul kanan hasil pemilah I juga akan dipilah agar terbentuk kelompok yang lebih homogen. Data anggota simpul kanan hasil pemilah I ditampilkan pada Tabel 2.8. Median dari data X_2 yaitu 50, kemudian dijadikan *threshold* pemilah simpul kanan. Sehingga, data X_2 yang ≤ 50 akan dipilah menjadi simpul kiri baru sedangkan yang > 50 dipilah menjadi simpul kanan baru. Tabel 2.9 menampilkan proporsi pemilahan simpul kanan menjadi simpul anak baru dan berikut ini perhitungan $i(t)$ yang dihasilkan.

Tabel 2.8 Data Anggota Simpul Kanan Hasil Pemilah I

No. Sampel Pengamatan	Y	X ₁	X ₂
1	3	1	50
5	3	1	60
6	2	2	50
8	3	1	50
10	3	1	50

Tabel 2.9 Proporsi Pengamatan Simpul Kanan Hasil Pemilah I jika Treshold $X_2 \leq 50$

		$N_L = 4$	$p_L = \frac{4}{5}$
Simpul kiri	Kelas 1	$N(1 t_L) = 0$	$p(1 t_L) = 0$
	Kelas 2	$N(2 t_L) = 1$	$p(2 t_L) = \frac{1}{4}$
	Kelas 3	$N(3 t_L) = 3$	$p(3 t_L) = \frac{3}{4}$
		$N_R = 1$	$p_R = \frac{1}{5}$
Simpul kanan	Kelas 1	$N(1 t_R) = 0$	$p(1 t_R) = 0$
	Kelas 2	$N(2 t_R) = 0$	$p(2 t_R) = 0$
	Kelas 3	$N(3 t_R) = 1$	$p(3 t_R) = 1$

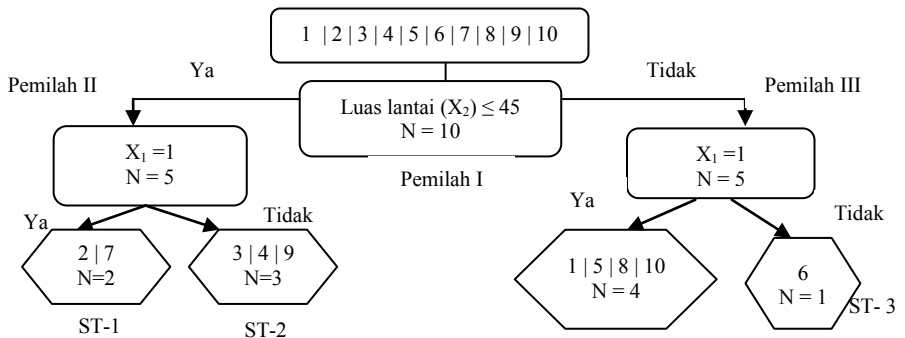
$$\begin{aligned}
 i(t) &= \frac{P_L P_R}{4} \left[\sum_j |p(j | t_L) - p(j | t_R)| \right]^2 \\
 &= \frac{(0.8)(0.2)}{4} [(0 - 0) + |\frac{1}{4} - 0| + |\frac{3}{4} - 1|]^2 \\
 &= 0.04 [0 + 0.25 + 0.25]^2 \\
 &= 0.01
 \end{aligned}$$

Tabel 2.10 Proporsi Pengamatan dari Simpul Kanan Hasil Pemilah I ke Simpul Kiri dan Kanan Jika Dipilah oleh $X_1=1$

		$N_L = 4$	$p_L = \frac{4}{5}$
Simpul kiri $X_1=1$	Kelas 1	$N(1 t_L)=0$	$p(1 t_L)=0$
	Kelas 2	$N(2 t_L)=0$	$p(2 t_L)=0$
	Kelas 3	$N(3 t_L)=4$	$p(3 t_L)=\frac{4}{5}$
		$N_R = 1$	$p_R = \frac{1}{5}$
Simpul kanan $X_1=2$	Kelas 1	$N(1 t_R)=0$	$p(1 t_R)=0$
	Kelas 2	$N(2 t_R)=1$	$p(2 t_R)=\frac{1}{5}$
	Kelas 3	$N(3 t_R)=0$	$p(3 t_R)=0$

$$\begin{aligned}
 i(t) &= \frac{p_L p_R}{4} \left[\sum_j |p(j|t_L) - p(j|t_R)| \right]^2 \\
 &= \frac{(0.8)(0.2)}{4} \left[|(0-0)| + |(0-\frac{1}{5})| + |(\frac{4}{5}-0)| \right]^2 \\
 &= 0.04 [0 + 0.2 + 0.8]^2 = 0.04
 \end{aligned}$$

Nilai $i(t)$ yang dihasilkan jika simpul kanan dipilah oleh $X_1=1$ lebih besar daripada jika dipilah $X_2 \leq 50$. Maka $X_1=1$ dipilih sebagai pemilah III, sehingga terbentuk pohon seperti Gambar 2.5.



Gambar 2.5 Contoh pohon klasifikasi ilustrasi I hasil pemilahan III.

Anggota simpul kanan baru tersebut hanya beranggotakan observasi ke-6 sehingga tidak bisa dipilah. Simpul tersebut dijadikan simpul terminal ke-3 (ST-3) yang digambarkan dengan bentuk segi enam. Sedangkan anggota simpul kiri baru memungkinkan dipilah. Median data diperoleh sebesar 50 sehingga pengamatan dengan $X_2 \leq 50$ akan dipilah menjadi anggota simpul kiri dan jika $X_2 > 50$ menjadi simpul kanan. Namun dari perhitungan $i(t)$ diperoleh sebesar 0 yang berarti sudah homogen. Maka simpul tersebut tidak perlu untuk dipilah dan akan dijadikan sebagai simpul terminal 4 (ST-4).

Tabel 2.11 Proporsi Pengamatan dari Simpul Kanan Hasil Pemilah III ke Simpul Kiri dan Kanan Jika Dipilah oleh $X_2 \leq 50$

		$N_L = 3$	$p_L = \frac{3}{4}$
Simpul kiri $X_1=1$	Kelas 1	$N(1 t_L)=0$	$p(1 t_L)=0$
	Kelas 2	$N(2 t_L)=0$	$p(2 t_L)=0$
	Kelas 3	$N(3 t_L)=3$	$p(3 t_L)=\frac{3}{3}=1$
		$N_R = 1$	$p_R = \frac{1}{4}$
Simpul kanan $X_1=2$	Kelas 1	$N(1 t_R)=0$	$p(1 t_R)=0$
	Kelas 2	$N(2 t_R)=0$	$p(2 t_R)=0$
	Kelas 3	$N(3 t_R)=1$	$p(3 t_R)=1$

$$\begin{aligned}
 i(t) &= \frac{p_L p_R}{4} \left[\sum_j |p(j|t_L) - p(j|t_R)| \right]^2 \\
 &= \frac{(0.75)(0.25)}{4} [(0-0) + |(0-0)| + |(1-1)|]^2 \\
 &= 0
 \end{aligned}$$

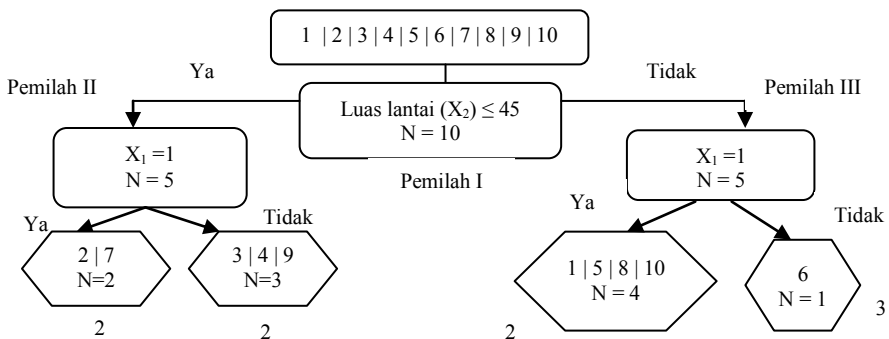
Dengan demikian pohon klasifikasi yang terbentuk memiliki empat simpul terminal. Langkah selanjutnya adalah melakukan penandaan label kelas dari masing-masing simpul terminal berdasarkan persamaan (2.5). Suatu simpul terminal bisa

jadi mempunyai label kelas yang sama dengan simpul terminal yang lainnya.

Tabel 2.12 Penentuan Penandaan Label Kelas pada Simpul Terminal

Simpul Terminal (ST)	Kelas	Jumlah	Persentase (%)	Label
ST-1	1	1	50	Kelas 2
	2	1	50	
	3	0	0	
	Total	2	100	
ST-2	1	1	33,3	Kelas 2
	2	2	66,7	
	3	0	0	
	Total	3	100	
ST-3	1	0	0	Kelas 2
	2	1	100	
	3	0	0	
	Total	1	100	
ST-4	1	0	0	Kelas 3
	2	0	0	
	3	3	100	
	Total	3	100	

Jadi, pohon klasifikasi yang terbentuk untuk data ilustrasi I yang diberikan dengan menggunakan aturan indeks Twoing ditampilkan Gambar 2.6.



Gambar 2.6 Contoh pohon klasifikasi hasil ilustrasi data I yang terbentuk.

2.1.6 Ilustrasi Pemilihan Pemilah Menggunakan Indeks Gini

Berikut ini diberikan ilustrasi lain pembentukan pohon klasifikasi dengan menggunakan fungsi keheterogenan indeks Gini, dimana variabel yang digunakan meliputi satu variabel respon dan tiga variabel prediktor. Masing-masing variabel tersebut secara lebih rinci ditampilkan dalam Tabel 2.13 sebagai berikut.

Tabel 2.13 Variabel untuk Ilustrasi II

	Variabel	Skala Data	Kategori atau Kelas
Y	Status Rumah Tangga	Ordinal	1. RT SM 2. RT M 3. RT HM
X ₁	Luas kavling termasuk bangunan (m ²)	Nominal	1. < 60 m ² 2. ≥ 60 m ²
X ₂	Luas lantai (m ²)	Nominal	1. < 32 m ² 2. ≥ 32 m ²
X ₃	Jenis dinding terluas	Nominal	1. Tembok 2. Kayu 3. Bambu

Ada 10 data pengamatan diberikan sebagai data ilustrasi yang akan diolah sesuai dengan algoritma CART. Data tersebut ditampilkan dalam Tabel 2.14.

Tabel 2.14 Data Ilustrasi II

No. Sampel Pengamatan	Y	X ₁	X ₂	X ₃
1	3	1	2	1
2	1	1	2	2
3	1	2	2	3
4	2	2	1	1
5	3	1	1	2
6	2	2	2	2
7	1	1	1	3
8	3	1	1	2
9	2	2	2	3
10	3	1	2	1

Pada awalnya, seluruh data sampel pengamatan tersebut berada dalam satu himpunan yang masih bersifat heterogen. Algoritma CART yang pertama yaitu pembentukan pohon klasifikasi, diawali dengan pemilihan pemilah untuk simpul utama. Pemilihan pemilah dimulai dengan menghitung indeks Gini. Lalu masing-masing variabel dan semua kemungkinan *threshold* atau nilai variabel yang menjadi pemilah dihitung nilai *goodness of split* dan yang menghasilkan *goodness of split* maksimum akan dipilih sebagai pemilah terbaik. Berikut ini masing-masing perhitungannya.

a. Untuk variabel Y

Tabel 2.15 Perhitungan Proporsi Pengamatan dari Y

Variabel	Kategori atau kelas	Jumlah Pengamatan	Proporsi
Y	1	3	$\frac{3}{10}$
	2	3	$\frac{3}{10}$
	3	4	$\frac{4}{10}$
Total		10	1

$$\begin{aligned}
 i(t) &= \sum_{i \neq j} p(i|t)p(j|t) \\
 &= [p(1|t) \cdot p(2|t)] + [p(1|t) \cdot p(3|t)] + [p(2|t) \cdot p(3|t)] \\
 &= \left[\frac{3}{10} \cdot \frac{3}{10}\right] + \left[\frac{3}{10} \cdot \frac{4}{10}\right] + \left[\frac{3}{10} \cdot \frac{4}{10}\right] \\
 &= 0.33
 \end{aligned}$$

b. Untuk variabel X_1

Variabel X_1 terdiri atas 2 level kategori berskala numerik sehingga kemungkinan pemilah ada sebanyak $2^{2-1} - 1 \approx 1$ kemungkinan yaitu pemilah $X_1=1$ sebagai simpul kiri dan $X_1=2$ sebagai simpul kanan. Dengan berdasarkan perhitungan proporsi pengamatan sebagaimana diberikan pada Tabel 2.16, selanjutnya bisa dihitung nilai $i(t)$ dan *goodness of split*.

Untuk simpul kiri:

$$\begin{aligned}
 i(t_L) &= \sum_{i \neq j} p(i|t_L)p(j|t_L) \\
 &= [p(1|t_L) \cdot p(2|t_L)] + [p(1|t_L) \cdot p(3|t_L)] + [p(2|t_L) \cdot p(3|t_L)] \\
 &= \left[\frac{2}{6} \cdot 0\right] + \left[\frac{2}{6} \cdot \frac{4}{6}\right] + \left[0 \cdot \frac{4}{6}\right] = 0.2222
 \end{aligned}$$

Tabel 2.16 Proporsi Pengamatan Data Ilustrasi II ke Simpul Kiri dan Kanan Jika Dipilah oleh X_1

		$N_L = 6$	$p_L = \frac{6}{10}$
Simpul kiri	Kelas 1	$N(1 t_L) = 2$	$p(1 t_L) = \frac{2}{6}$
	Kelas 2	$N(2 t_L) = 0$	$p(2 t_L) = 0$
	Kelas 3	$N(3 t_L) = 4$	$p(3 t_L) = \frac{4}{6}$
		$N_R = 4$	$p_R = \frac{4}{10}$
Simpul kanan	Kelas 1	$N(1 t_R) = 1$	$p(1 t_R) = \frac{1}{4}$
	Kelas 2	$N(2 t_R) = 3$	$p(2 t_R) = \frac{3}{4}$
	Kelas 3	$N(3 t_R) = 0$	$p(3 t_R) = 0$

Untuk simpul kanan:

$$\begin{aligned}
 i(t_R) &= \sum_{i \neq j} p(i|t_R)p(j|t_R) \\
 &= [p(1|t_R) \cdot p(2|t_R)] + [p(1|t_R) \cdot p(3|t_R)] + [p(2|t_R) \cdot p(3|t_R)] \\
 &= [\frac{1}{4} \cdot \frac{3}{4}] + [\frac{1}{4} \cdot 0] + [\frac{3}{4} \cdot 0] \\
 &= 0.1875
 \end{aligned}$$

Maka dengan persamaan (2.3) dapat dihitung nilai *goodness of split* untuk pemilah X_1 tersebut sebagai berikut.

$$\begin{aligned}
 \phi(s, t) &= \Delta i(s, t) = i(t) - p_L i(t_L) - p_R i(t_R) \\
 &= 0.33 - (\frac{6}{10} \cdot 0.2222) - (\frac{4}{10} \cdot 0.1875) = 0.1217
 \end{aligned}$$

c. Untuk variabel X_2

Variabel X_2 terdiri atas 2 level kategori berskala numerik sehingga kemungkinan pemilah sama seperti pada X_1 yaitu ada 1 kemungkinan dimana pemilah $X_2=1$ sebagai simpul kiri dan $X_2=2$ sebagai simpul kanan.

Untuk simpul kiri:

$$\begin{aligned}
 i(t_L) &= \sum_{i \neq j} p(i|t_L)p(j|t_L) \\
 &= [p(1|t_L) \cdot p(2|t_L)] + [p(1|t_L) \cdot p(3|t_L)] + [p(2|t_L) \cdot p(3|t_L)]
 \end{aligned}$$

$$\begin{aligned}
&= \left[\frac{1}{4} \cdot \frac{1}{4} \right] + \left[\frac{1}{4} \cdot \frac{2}{4} \right] + \left[\frac{1}{4} \cdot \frac{2}{4} \right] \\
&= 0.3125
\end{aligned}$$

Tabel 2.17 Proporsi Pengamatan Data Ilustrasi II ke Simpul Kiri dan Kanan jika Dipilah oleh X_2

		$N_L = 4$	$p_L = \frac{4}{10}$
Simpul kiri	Kelas 1	$N(1 t_L) = 1$	$p(1 t_L) = \frac{1}{4}$
	Kelas 2	$N(2 t_L) = 1$	$p(2 t_L) = \frac{1}{4}$
	Kelas 3	$N(3 t_L) = 2$	$p(3 t_L) = \frac{2}{4}$
		$N_R = 6$	$p_R = \frac{6}{10}$
Simpul kanan	Kelas 1	$N(1 t_R) = 2$	$p(1 t_R) = \frac{2}{6}$
	Kelas 2	$N(2 t_R) = 2$	$p(2 t_R) = \frac{2}{6}$
	Kelas 3	$N(3 t_R) = 2$	$p(3 t_R) = \frac{2}{6}$

Untuk simpul kanan:

$$\begin{aligned}
i(t_R) &= \sum_{i \neq j} p(i | t_R) p(j | t_R) \\
&= [p(1 | t_R) \cdot p(2 | t_R)] + [p(1 | t_R) \cdot p(3 | t_R)] + [p(2 | t_R) \cdot p(3 | t_R)] \\
&= \left[\frac{2}{6} \cdot \frac{2}{6} \right] + \left[\frac{2}{6} \cdot \frac{2}{6} \right] + \left[\frac{2}{6} \cdot \frac{2}{6} \right] = 0.3333
\end{aligned}$$

Maka, nilai *goodness of split* untuk pemilah X_2 tersebut sebagai berikut.

$$\begin{aligned}
\phi(s, t) &= \Delta i(s, t) = i(t) - p_L i(t_L) - p_R i(t_R) \\
&= 0.33 - \left(\frac{4}{10} \cdot 0.3125 \right) - \left(\frac{6}{10} \cdot 0.333 \right) \\
&= 0.0052
\end{aligned}$$

d. Untuk variabel X_3

Variabel X_3 terdiri atas 3 level kategori berskala numerik sehingga kemungkinan pemilah ada sebanyak $2^{3-1}-1 \approx 3$ kemungkinan yang ditampilkan pada Tabel 2.18. Telah dijelaskan sebelumnya bahwa metode CART mengikuti prosedur *binary recursive partitioning*. Maka, pemilah hanya bisa memilah variabel prediktor pada suatu simpul induk menjadi 2 simpul anak yaitu simpul kiri dan simpul kanan meskipun variabel prediktor tersebut mempunyai kategori lebih dari 2.

Tabel 2.18 Kemungkinan Pemilah X_3

Kemungkinan ke-	Simpul kiri	Simpul kanan
I	$X_3 = 1$	$X_3 = 2 \text{ \& } 3$
II	$X_3 = 2$	$X_3 = 1 \text{ \& } 3$
III	$X_3 = 3$	$X_3 = 1 \text{ \& } 2$

Pada kemungkinan I, variabel X_3 dengan *threshold* 1 dengan 2 & 3 digunakan sebagai pemilah. X_3 yang termasuk kategori 1 (ada 3 pengamatan) akan dipilah menjadi simpul kiri, sedangkan yang termasuk kategori 2 (ada 4 pengamatan) dan kategori 3 (ada 3 pengamatan) akan menjadi simpul kanan. Dengan demikian diperoleh jumlah pengamatan pada simpul kiri atau $N_L = 3$ sementara jumlah pengamatan pada simpul kanan atau $N_R = 3 + 4 = 7$. Selanjutnya dapat dihitung nilai *goodness of split* sebagai berikut.

Tabel 2.19 Proporsi Pengamatan jika Dipilah oleh X_3 (Kemungkinan I)

		$N_L = 3$	$p_L = \frac{3}{10}$
Simpul kiri	Kelas 1	$N(1 t_L) = 0$	$p(1 t_L) = 0$
	Kelas 2	$N(2 t_L) = 1$	$p(2 t_L) = \frac{1}{3}$
	Kelas 3	$N(3 t_L) = 2$	$p(3 t_L) = \frac{2}{3}$

Tabel 2.19 Lanjutan

		$N_R = 7$	$p_R = \frac{7}{10}$
	Kelas 1	$N(1 t_R) = 3$	$p(1 t_R) = \frac{3}{7}$
Simpul kanan	Kelas 2	$N(2 t_R) = 2$	$p(2 t_R) = \frac{2}{7}$
	Kelas 3	$N(3 t_R) = 2$	$p(3 t_R) = \frac{2}{7}$

Untuk simpul kiri:

$$\begin{aligned}
 i(t_L) &= \sum_{i \neq j} p(i | t_L) p(j | t_L) \\
 &= [p(1 | t_L) \cdot p(2 | t_L)] + [p(1 | t_L) \cdot p(3 | t_L)] + [p(2 | t_L) \cdot p(3 | t_L)] \\
 &= \left[0 \cdot \frac{1}{3}\right] + \left[0 \cdot \frac{2}{3}\right] + \left[\frac{1}{3} \cdot \frac{2}{3}\right] \\
 &= 0.2222
 \end{aligned}$$

Untuk simpul kanan:

$$\begin{aligned}
 i(t_R) &= \sum_{i \neq j} p(i | t_R) p(j | t_R) \\
 &= [p(1 | t_R) \cdot p(2 | t_R)] + [p(1 | t_R) \cdot p(3 | t_R)] + [p(2 | t_R) \cdot p(3 | t_R)] \\
 &= \left[\frac{3}{7} \cdot \frac{2}{7}\right] + \left[\frac{3}{7} \cdot \frac{2}{7}\right] + \left[\frac{2}{7} \cdot \frac{2}{7}\right] \\
 &= 0.3265
 \end{aligned}$$

Maka, nilai *goodness of split* untuk pemilah X_3 kemungkinan I sebagai berikut.

$$\begin{aligned}
 \phi(s, t) &= \Delta i(s, t) = i(t) - p_L i(t_L) - p_R i(t_R) \\
 &= 0.33 - \left(\frac{3}{10} \cdot 0.2222\right) - \left(\frac{7}{10} \cdot 0.3265\right) \\
 &= 0.0349
 \end{aligned}$$

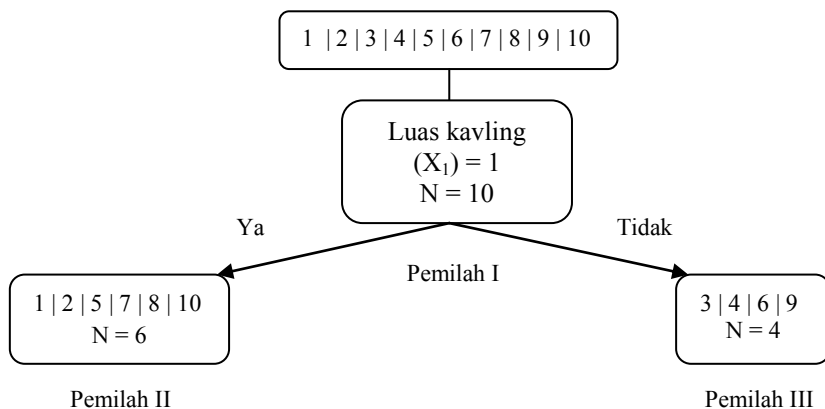
Dengan cara yang sama, maka dapat dihitung nilai *goodness of split* untuk pemilah X_3 kemungkinan II dan kemungkinan III. Selanjutnya perhitungan *goodness of split* masing-masing dapat dilihat rangkuman hasilnya pada Tabel 2.20.

Tabel 2.20 *Goodness of Split* Kemungkinan Pemilah I

Variabel	Threshold	<i>Goodness of Split</i>
X_1	$X_1 = 1 \mid X_1 = 2$	0.1217 ^{*)}
X_2	$X_2 = 1 \mid X_2 = 2$	0.0052
X_3	$X_3 = 1 \mid X_3 = 2 \ \& \ 3$	0.0349
	$X_3 = 2 \mid X_3 = 1 \ \& \ 3$	0.0052
	$X_3 = 3 \mid X_3 = 1 \ \& \ 2$	0.0633

^{*)} = Nilai *goodness of split* maksimum

Berdasarkan kriteria nilai *Goodness of Split* yang maksimum, maka X_1 yaitu luas kavling termasuk bangunan (m^2) dipilih sebagai variabel prediktor yang menjadi simpul utama (pemilah I). Nilai variabel X_1 yang menjadi pemilah ditentukan menurut nilai peluang antara kelas $i = 1$ atau $p(i=1|t)$ dengan kelas $j = 2$ atau $p(j=2|t)$ yang lebih besar. Diketahui bahwa nilai peluang untuk $X_1 = 1$ yaitu 0.6 lebih besar daripada peluang $X_1 = 2$ yang hanya 0.4 (lihat Tabel 2.16). Maka $X_1 = 1$ dijadikan sebagai *threshold* pemilah simpul utama menjadi simpul anak kiri dan kanan. Dengan demikian data sampel pengamatan selanjutnya dipilih berdasarkan pemilah I tersebut yaitu yang termasuk ke dalam simpul kiri adalah data sampel dengan X_1 yang berkelas 1 ($X_1 = 1$) sedangkan data sampel X_1 yang berkelas 2 ($X_1 = 2$) termasuk ke dalam simpul kanan. Pohon klasifikasi awal yang dihasilkan sebagai berikut.

**Gambar 2.7** Contoh pohon klasifikasi ilustrasi II hasil pemilahan I.

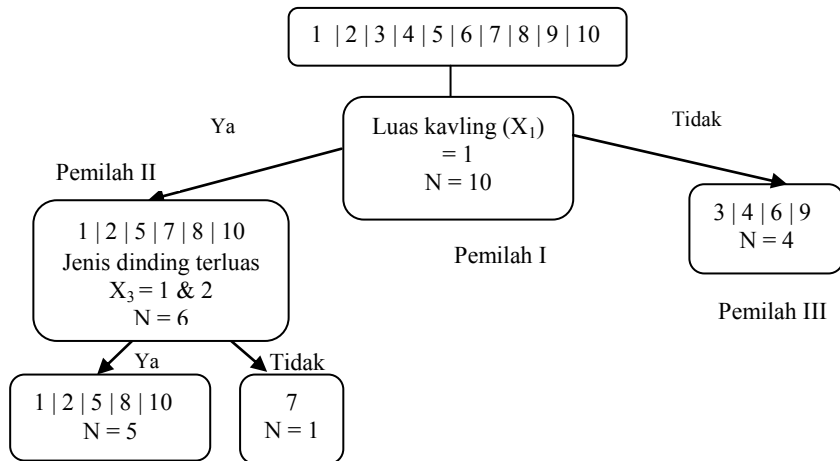
Simpul kiri dan simpul kanan yang dihasilkan dari pemilahan I masih bersifat heterogen sehingga perlu dilakukan pemilahan kembali. Pemilah II akan memilah sampel pengamatan dalam simpul kiri $X_1 = 1$ menjadi simpul anak sementara pemilah III akan memilah sampel pengamatan dalam simpul kiri $X_1 = 2$ menjadi simpul anak. Cara mendapatkannya sama, yaitu menghitung nilai indeks Gini dan *goodness of split* untuk semua variabel dan nilai variabel yang mungkin lalu dipilih yang menghasilkan nilai *goodness of split* maksimum. Dari perhitungan nilai indeks Gini dan *goodness of split* untuk semua variabel dan nilai variabel yang mungkin, diperoleh hasil pada Tabel 2.21. Berdasarkan kriteria nilai *Goodness of Split* yang maksimum, maka X_3 yaitu jenis dinding terluas dipilih sebagai pemilah II.

Tabel 2.21 Nilai *Goodness of Split* Kemungkinan Pemilah II

Variabel	Threshold	<i>Goodness of Split</i>
X_2	$X_2 = 1 \mid X_2 = 2$	0.000
X_3	$X_3 = 1 \mid X_3 = 2 \ \& \ 3$	0.194
	$X_3 = 2 \mid X_3 = 1 \ \& \ 3$	0.222
	$X_3 = 3 \mid X_3 = 1 \ \& \ 2$	0.3107 ^{*)}

^{*)} = Nilai *goodness of split* maksimum

Nilai variabel X_3 yang menjadi pemilah ditentukan menurut nilai peluang antara kelas $i = 3$ atau $p(i = 3 \mid t)$ dengan kelas $j=1\&2$ atau $p(j = 1;2 \mid t)$ yang lebih besar. Hasil perhitungan menunjukkan bahwa nilai peluang untuk $X_3 = 3$ sebesar 0.166, lebih kecil daripada peluang $X_3 = 1 \ \& \ 2$ yang mencapai 0.833. Maka $X_3 = 1 \ \& \ 2$ dijadikan sebagai *threshold* pemilah II. Jadi, data sampel pengamatan yang awalnya termasuk dalam simpul kiri hasil pemilah I selanjutnya dipilih berdasarkan pemilah II tersebut yaitu yang termasuk ke dalam simpul anak kiri yang baru adalah data sampel X_1 yang berkelas 1 ($X_1 = 1$) dan X_3 yang berkelas 1 & 2 ($X_3 = 1 \ \& \ 2$) sedangkan data sampel X_1 yang berkelas 1 ($X_1 = 1$) dan X_3 berkelas 3 ($X_3 = 3$) termasuk ke dalam simpul kanan yang baru. Pohon klasifikasi awal yang dihasilkan sebagai berikut.



Gambar 2.8 Contoh pohon klasifikasi ilustrasi II hasil pemilahan II.

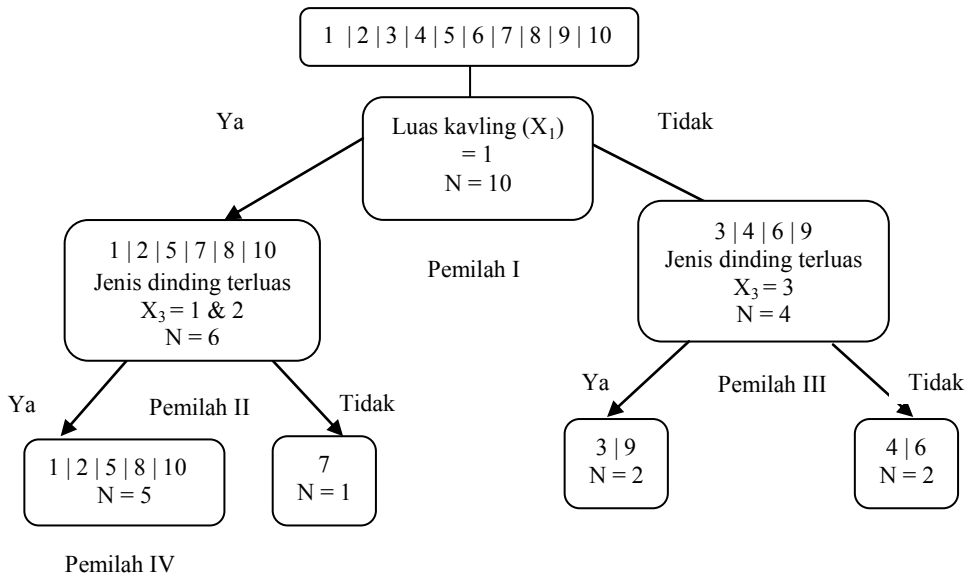
Sementara hasil perhitungan nilai indeks Gini dan *goodness of split* untuk semua variabel dan nilai variabel yang mungkin sebagai pemilah III ditampilkan dalam Tabel 2.22.

Tabel 2.22 *Goodness of Split* Kemungkinan Pemilah III

Variabel	Threshold	<i>Goodness of Split</i>
X_2	$X_2 = 1 \mid X_2 = 2$	0.2085
X_3	$X_3 = 1 \mid X_3 = 2 \ \& \ 3$	0.2085
	$X_3 = 2 \mid X_3 = 1 \ \& \ 3$	0.2085
	$X_3 = 3 \mid X_3 = 1 \ \& \ 2$	0.25 ^{*)}

^{*)} = Nilai *goodness of split* maksimum

Berdasarkan kriteria nilai *Goodness of Split* yang maksimum, X_3 (jenis dinding terluas) dipilih sebagai pemilah III. Nilai variabel X_3 yang menjadi pemilah ditentukan menurut nilai peluang yang lebih besar antara kelas $i = 3$ atau $p(i = 3 \mid t)$ dengan kelas $j = 1 \ \& \ 2$ atau $p(j = 1; 2 \mid t)$. Hasil perhitungan menunjukkan bahwa nilai peluang untuk $X_3 = 3$ maupun peluang $X_3 = 1 \ \& \ 2$ sama-sama bernilai 0.5. Maka dalam hal ini peneliti memilih $X_3 = 3$ dijadikan sebagai *threshold* pemilah III. Maka data sampel pengamatan dalam simpul kanan hasil pemilah I selanjutnya dipilah berdasarkan pemilah III tersebut



Gambar 2.9 Contoh pohon klasifikasi ilustrasi II hasil pemilahan III.

Gambar 2.9 menunjukkan pohon klasifikasi setelah dilakukan pemilahan III dimana, yang termasuk ke dalam simpul anak kiri yang baru adalah data sampel X_1 berkelas 2 ($X_1 = 2$) dan X_3 yang berkelas 3 ($X_3 = 3$). Sedangkan data sampel X_1 yang berkelas 2 ($X_1 = 2$) dan X_3 berkelas 1 & 2 ($X_3 = 1 \& 2$) termasuk ke dalam simpul kanan yang baru.

Simpul anak kiri hasil pemilah II masih bisa dipilah lagi agar lebih homogen sedangkan simpul kanannya dianggap menjadi simpul terminal karena hanya mempunyai 1 anggota sampel pengamatan. Dengan pemilah IV, simpul anak kiri hasil pemilah II akan dipilah lagi menjadi simpul anak kiri dan kanan yang diharapkan akan lebih homogen. Berikut adalah hasil perhitungan nilai indeks Gini dan *goodness of split* untuk semua variabel dan nilai variabel yang mungkin sebagai pemilah IV.

Tabel 2.23 Nilai *Goodness of Split* Kemungkinan Pemilah IV

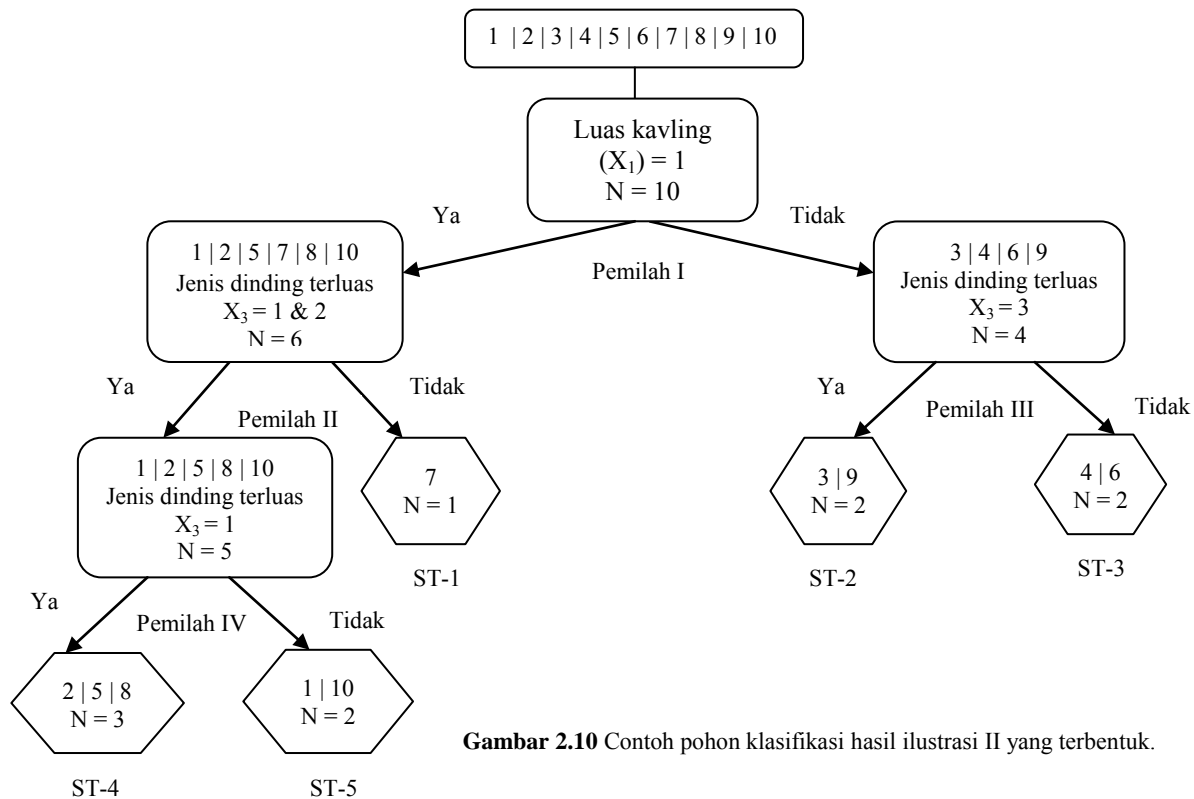
Variabel	Threshold	Goodness of Split
X_2	$X_2 = 1 \mid X_2 = 2$	0.0266
X_3	$X_3 = 1 \mid X_3 = 2$	0.0266

Berdasarkan Tabel 2.23 diketahui bahwa *goodness of split* dari kedua pemilah yang mungkin mempunyai nilai yang sama yaitu 0.0266. Jika dilihat datanya yang ditampilkan Tabel 2.24, maka tampak bahwa kedua kemungkinan pemilah tersebut memberikan hasil pemilahan yang sama. Peneliti harus memilih salah satu dari pemilah tersebut.

Tabel 2.24 Data Anggota Simpul Kiri Hasil Pemilah II

No. Sampel Pengamatan	Y	X ₁	X ₂	X ₃
1	3	1	2	1
2	1	1	2	2
5	3	1	1	2
8	3	1	1	2
10	3	1	2	1

Peneliti memilih variabel X₃ dengan pertimbangan simpul kiri yang menjadi simpul induk pemilah IV merupakan data sampel pengamatan dengan X₃ kelas 1&2. Nilai variabel X₃ yang menjadi pemilah ditentukan menurut nilai peluang yang lebih besar antara kelas $i=1$ atau $p(i=1|t)$ dengan kelas $j=2$ atau $p(j=2|t)$. Hasil perhitungan menunjukkan bahwa nilai peluang untuk X₃ = 1 sebesar 0.4, lebih kecil daripada peluang X₃ = 2 yang mencapai 0.6. Maka X₃ = 2 dijadikan sebagai *threshold* pemilah IV menjadi simpul anak kiri dan kanan. Dengan demikian data sampel pengamatan yang awalnya termasuk dalam simpul kiri hasil pemilah II selanjutnya dipilah berdasarkan pemilah IV tersebut yaitu yang termasuk ke dalam simpul anak kiri yang baru adalah data sampel X₁ yang berkelas 1 (X₁ = 1) dan X₃ yang berkelas 2 (X₃ = 2) sedangkan data sampel X₁ yang berkelas 1 (X₁ = 1) dan X₃ berkelas 1 (X₃ = 1) termasuk ke dalam simpul kanan yang baru. Sementara itu, simpul kiri dan simpul kanan hasil pemilah III dianggap sudah homogen dan dijadikan simpul terminal karena anggota sampel pengamatan di dalamnya hanya terdiri atas 2 sampel pengamatan. Pohon klasifikasi yang dihasilkan tampak sebagai berikut.



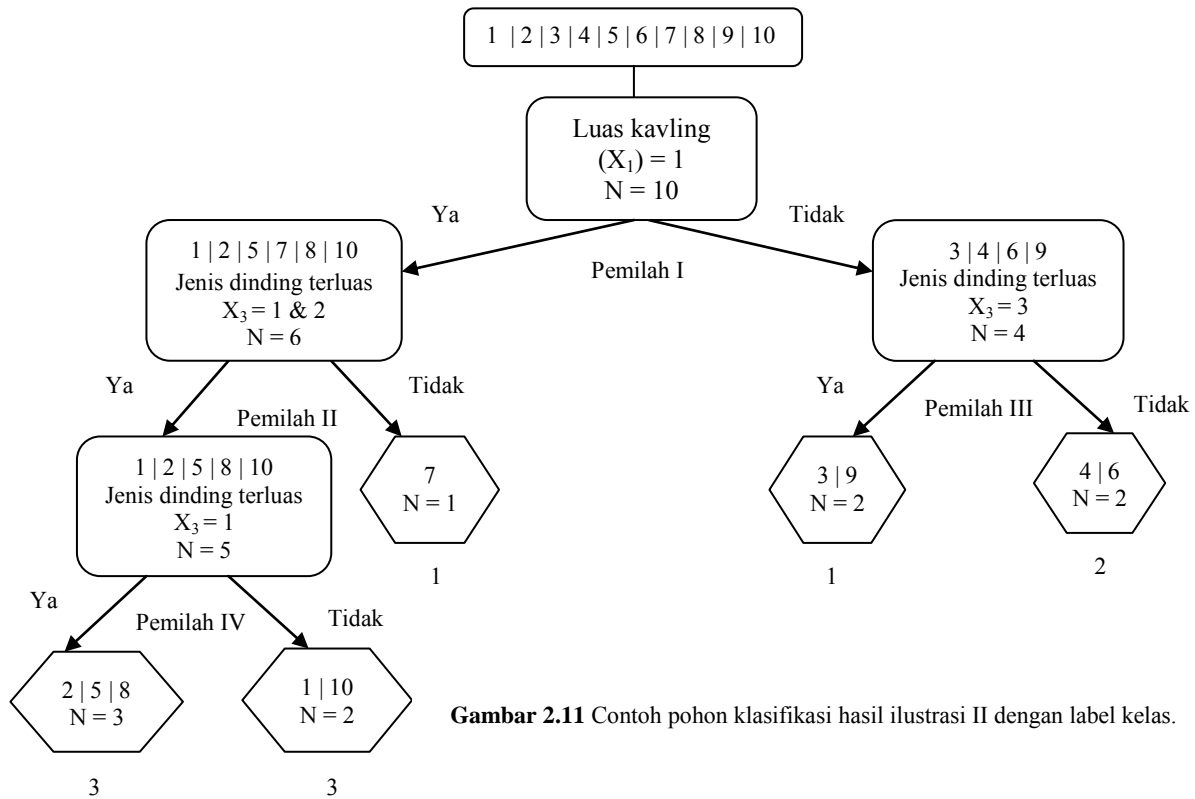
Gambar 2.10 Contoh pohon klasifikasi hasil ilustrasi II yang terbentuk.

Pohon klasifikasi yang terbentuk mempunyai lima simpul terminal yang pada Gambar 3.4 ditunjukkan oleh ST-1, ST-2, ST-3, ST-4 dan ST-5. Langkah selanjutnya adalah melakukan penandaan label kelas dari masing-masing simpul terminal.

Tabel 2.25 Penentuan Penandaan Label Kelas Simpul Terminal

Simpul Terminal (ST)	Kelas	Jumlah	Persentase (%)	Label
ST-1	1	1	100	Kelas 1
	2	0	0	
	3	0	0	
	Total	1	100	
ST-2	1	1	50	Kelas 1
	2	1	50	
	3	0	0	
	Total	2	100	
ST-3	1	0	0	Kelas 2
	2	2	100	
	3	0	0	
	Total	2	100	
ST-4	1	1	33.3	Kelas 3
	2	0	0	
	3	2	66.7	
	Total	3	100	
ST-5	1	0	0	Kelas 3
	2	0	0	
	3	2	100	
	Total	2	100	

Jadi, pohon klasifikasi yang terbentuk untuk data ilutrasi II yang diberikan adalah sebagai berikut.



Gambar 2.11 Contoh pohon klasifikasi hasil ilustrasi II dengan label kelas.

2.2 Teori Kemiskinan

Mengacu pada program Strategi Nasional Penanggulangan Kemiskinan, kemiskinan didefinisikan sebagai suatu kondisi dimana seseorang atau sekelompok orang, laki-laki dan perempuan tidak terpenuhi hak-hak dasarnya untuk mempertahankan dan mengembangkan kehidupan yang bermartabat. Definisi ini beranjak dari pendekatan berbasis hak yang mengakui bahwa masyarakat miskin mempunyai hak-hak dasar yang sama dengan anggota masyarakat lainnya.

Badan Pusat Statistik atau BPS (2008) mendefinisikan Garis Kemiskinan (GK) sebagai nilai rupiah yang harus dikeluarkan seseorang dalam sebulan agar dapat memenuhi kebutuhan dasar asupan kalori sebesar 2100 kkal/hari per kapita (Garis Kemiskinan Makanan atau GKM) ditambah kebutuhan minimum non makanan yang merupakan kebutuhan seseorang yaitu papan, sandang, sekolah, transportasi dan kebutuhan individu rumah tangga dasar lainnya (Garis Kemiskinan Non Makanan atau GKNM). *World Bank Institute* mendefinisikan kemiskinan absolut adalah hidup dengan pendapatan USD \$ 1 per hari. Kemiskinan relatif merupakan kondisi miskin karena kebijakan pembangunan yang belum mampu menjangkau seluruh lapisan masyarakat. BPS menggunakan tiga indikator kemiskinan yaitu:

1. *Head Count Index* (HCI- P_0) yaitu persentase penduduk yang berada di bawah Garis Kemiskinan (GK).
2. Indeks Kedalaman Kemiskinan (*Poverty Gap Index*- P_1) yang merupakan ukuran rata-rata kesenjangan penyebaran pengeluaran masing-masing penduduk miskin terhadap garis kemiskinan. Semakin tinggi nilai indeks, semakin jauh rata-rata pengeluaran penduduk dari garis kemiskinan.
3. Indeks Keparahan Kemiskinan (*Poverty Severity Indeks*- P_2) yang memberikan gambaran mengenai penyebaran pengeluaran diantara penduduk miskin. Semakin tinggi nilai indeks, semakin tinggi ketimpangan pengeluaran diantara penduduk miskin.

Untuk memenuhi kebutuhan data rumah tangga, BPS telah melakukan pendataan rumah tangga baik dalam Sensus Penduduk, Supas maupun Susenas. Bahkan pada akhir tahun

2005 telah dilakukan pendataan khusus rumah tangga miskin dengan menggunakan 14 indikator kemiskinan untuk memenuhi kebutuhan berbagai program pelayanan dasar tersebut yang meliputi:

- 1) Luas lantai rumah
- 2) Jenis lantai rumah
- 3) Jenis dinding rumah
- 4) Fasilitas tempat buang air besar
- 5) Sumber air minum
- 6) Penerangan yang digunakan
- 7) Bahan bakar yang digunakan
- 8) Frekuensi makan dalam sehari
- 9) Kebiasaan membeli daging/ayam/susu
- 10) Kemampuan membeli pakaian
- 11) Kemampuan berobat ke puskesmas/poliklinik
- 12) Lapangan pekerjaan kepala rumah tangga
- 13) Pendidikan kepala rumah tangga
- 14) Kepemilikan aset

Dalam laporan SPDKP07 (Survei Pelayanan Dasar Kesehatan dan Pendidikan 2007) disebutkan beberapa kriteria umum RTSM (Rumah Tangga Sangat Miskin), yaitu:

- i. Sebagian besar pengeluarannya digunakan untuk memenuhi konsumsi makanan pokok yang sangat sederhana.
- ii. Biasanya tidak mampu atau mengalami kesulitan untuk berobat ke tenaga medis kecuali Puskesmas atau yang disubsidi pemerintah.
- iii. Tidak mampu membeli pakaian satu kali dalam satu tahun untuk setiap anggota rumah tangga.
- iv. Biasanya tidak/hanya mampu menyekolahkan anaknya sampai jenjang pendidikan SLTP.

Dari sisi kondisi fisik serta fasilitas tempat tinggal RTSM biasanya tinggal pada rumah yang:

- i. Dinding rumahnya terbuat dari bambu/kayu/tembok dengan kondisi tidak baik/kualitas rendah, termasuk tembok yang sudah usang/berlumut atau tembok tidak dipelster.

- ii. Sebagian besar lantai terbuat dari tanah atau kayu/semen/keramik dengan kondisi tidak baik/kualitas rendah.
- iii. Atap terbuat dari ijuk/rumbia atau genteng/seng/asbes dengan kondisi tidak baik/ kualitas rendah.
- iv. Penerangan bangunan tempat tinggal bukan dari listrik atau listrik tanpa meteran.
- v. Luas lantai rumah kecil (biasanya kurang dari 8 m²/orang).
- vi. Sumber air minum berasal dari sumur atau mata air tak terlindung/air sungai/air hujan/lainnya.

BPS mengadakan pemutakhiran (*updating*) indikator pada tahun 2011 dalam rangka mendukung Program Perlindungan Sosial 2011 (BPS PPLS, 2011).

2.3 Penelitian Sebelumnya

Ada dua penelitian yang dijadikan sebagai dasar acuan tentang penelitian dengan pendekatan metode CART ini. Penelitian yang dilakukan oleh Muttaqin pada tahun 2013 yaitu analisis metode *ensemble* CART untuk perbaikan klasifikasi kemiskinan di kabupaten Jombang dengan batasan penelitian yaitu data yang dipakai hanya menggunakan data rumah tangga miskin di kecamatan Diwek. Variabel penelitian yang diteliti meliputi variabel respon kategorik yang biner yaitu status rumah tangga miskin di kabupaten Jombang (rumah tangga miskin dan rumah tangga sangat miskin) serta variabel prediktor sebanyak 18 variabel yang diacu dari BPS. Muttaqin pada awalnya melakukan analisis klasifikasi menggunakan metode CART kemudian dilanjutkan dengan metode *ensemble*. Kesimpulan yang dihasilkan yaitu variabel yang digunakan sebagai pemilah pohon klasifikasi CART dan paling menentukan status kemiskinan rumah tangga secara berurutan adalah penghasilan kepala rumah tangga tiap bulan, luas lantai bangunan tempat tinggal, luas kavling termasuk bangunan, dan sumber air minum. Berkaitan dengan akurasi klasifikasi, data sampel yang tepat diklasifikasikan secara keseluruhan sebesar 69,86 persen. Namun, akurasi prediksi pohon klasifikasi CART untuk kelas sangat miskin hanya sebesar 5,02 persen.

Selain itu, Paramita pada tahun 2010 pernah melakukan klasifikasi sifat curah hujan berdasarkan indikator ENSO (*El-Nino Southern Oscillation*) di kabupaten Ngawi dengan metode klasifikasi pohon dimana variabel respon terdiri atas 3 kategori. Dalam penelitian tersebut, lokasi stasiun (pos hujan) dibagi menjadi 2 Zona Musim (ZOM) yaitu ZOM-1 dan ZOM-2. Masing-masing dicari model pohon klasifikasi yang terbaik. Variabel penelitian meliputi 8 variabel prediktor dan 1 variabel respon yaitu rata-rata curah hujan bulanan dalam mm (milimeter) yang dikategorikan menjadi tiga; curah hujan di bawah normal, curah hujan normal, dan curah hujan di atas normal. Data yang digunakan sebanyak 240 pengamatan, yang mana pada saat pembentukan pohon klasifikasi, data tersebut dibagi menjadi data *learning* sebanyak 95 persen (228 pengamatan) dan sisanya sebagai data *testing* (5 persen atau 12 pengamatan). Hasil penelitian menunjukkan bahwa variabel prediktor yang paling berpengaruh terhadap curah hujan bulanan di ZOM-1 dan ZOM-2 adalah variabel *Dipole Mode Index* (DMI). Ketepatan klasifikasi untuk data *learning* dan *testing* yang dihasilkan oleh pohon optimal adalah sebesar 49,60 persen dan 50 persen untuk ZOM-1 sementara untuk ZOM-2 yaitu 59,60 persen dan 41,70 persen.

Penelitian mengenai kemiskinan di kabupaten Jombang yang dijadikan sebagai dasar acuan tentang penelitian ini, antara lain penelitian Susiani tahun 2013 tentang penentuan indikator kemiskinan berdasarkan dimensi kualitas kesehatan dan kualitas ekonomi menggunakan *Confirmatory Factor Analysis* (CFA) dengan pendekatan Bayesian. Diperoleh kesimpulan bahwa fasilitas pengobatan memberikan pengaruh terbesar dalam mengukur kualitas kesehatan sementara indikator kepemilikan bangunan memberikan pengaruh terbesar dalam mengukur kualitas ekonomi. Selain itu, Ningrum (2013) mengidentifikasi keterkaitan antar faktor penentuan kemiskinan di Kabupaten Jombang dengan menggunakan metode SEM dan diperoleh kesimpulan bahwa ekonomi dan Sumber Daya Manusia (SDM) berpengaruh positif terhadap kemiskinan dengan asumsi melihat indikator-indikator yang membentuk laten SDM dan ekonomi, sedangkan kesehatan berpengaruh negatif terhadap kemiskinan.

(Halaman ini sengaja dikosongkan)

BAB III METODE PENELITIAN

3.1 Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder dari hasil Survei Verifikasi Rumah Tangga Miskin Kabupaten Jombang yang dirancang oleh Badan Perencanaan Pembangunan (Bappeda) kabupaten Jombang tahun 2010. Unit penelitian dalam penelitian ini adalah rumah tangga miskin di Kabupaten Jombang yang jumlahnya mencapai 73.720 rumah tangga miskin. Tim survei verifikasi RTM kabupaten Jombang dipersiapkan dan dibagi untuk mendatangi para rumah tangga miskin yang telah terdata lalu melakukan pendataan verifikasi kemiskinan dengan membagikan kuesioner-kuesioner yang sudah disiapkan untuk diisi oleh responden.

3.2 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini meliputi satu variabel respon dengan tiga level kategorik dan ada delapan belas variabel bebas yang mengacu pada penelitian BPS sebagaimana diuraikan pada subbab 2.2. Masing-masing variabel tersebut dirinci dalam Tabel 3.1.

Tabel 3.1 Variabel Penelitian

	Variabel	Skala Data	Kategori
Y	Status kemiskinan rumah tangga miskin menurut BPS	Ordinal	1. RT Sangat Miskin (RTSM) 2. RT Miskin (RTM) 3. RT Hampir Miskin (RTHM)
X ₁	Status penguasaan bangunan tempat tinggal	Nominal	1. Milik sendiri 2. Kontrak 3. Sewa 4. Bebas sewa 5. Rumah dinas 6. Rumah milik orang tua/ sanak saudara
X ₂	Luas kavling termasuk bangunan (m ²)	Rasio	-

Tabel 3.1 Variabel Penelitian Lanjutan

	Variabel	Skala Data	Kategori
X ₃	Luas lantai (m ²)	Rasio	-
X ₄	Jenis atap rumah terluas	Nominal	<ol style="list-style-type: none"> 1. Beton 2. Genteng 3. Kayu Sirap 4. Seng 5. Asbes 6. Ijuk/rumbia
X ₅	Jenis dinding terluas	Nominal	<ol style="list-style-type: none"> 1. Tembok 2. Kayu 3. Bambu
X ₆	Jenis lantai terluas	Nominal	<ol style="list-style-type: none"> 1. Keramik/ marmer 2. Ubin/ tegel 3. Semen/ bata merah 4. Kayu/papan 5. Bambu 6. Tanah
X ₇	Fasilitas tempat buang air besar (jamban)	Nominal	<ol style="list-style-type: none"> 1. Milik sendiri 2. Milik bersama 3. Umum 4. Tidak ada
X ₈	Tempat pembuangan akhir tinja	Nominal	<ol style="list-style-type: none"> 1. Septictank 2. Kolam/sawah 3. Sungai/waduk 4. Lubang tanah 5. Tanah lapang/ kebun
X ₉	Sumber penerangan utama	Nominal	<ol style="list-style-type: none"> 1. Listrik PLN meteran 2. Listrik PLN bukan meteran (menumpang, dsb) 3. Listrik Non PLN 4. Bukan Listrik
X ₁₀	Sumber air minum	Nominal	<ol style="list-style-type: none"> 1. Air dalam kemasan 2. Ledeng 3. Pompa 4. Sumur 5. Mata air 6. Air sungai

Tabel 3.1 Variabel Penelitian Lanjutan

	Variabel	Skala Data	Kategori
X ₁₁	Bahan bakar memasak	Nominal	1. Listrik 2. Gas/elpiji 3. Minyak tanah 4. Arang kayu/ tempurung 5. Kayu bakar
X ₁₂	Intensitas konsumsi daging/ susu/ ayam per minggu	Rasio	-
X ₁₃	Intensitas membeli pakaian per tahun	Rasio	-
X ₁₄	Intensitas makan per hari	Rasio	-
X ₁₅	Pengobatan	Nominal	1. RS/ Puskesmas/ Pustu 2. Praktik dokter 3. Praktik paramedis 4. Praktik pengobatan tradisional
X ₁₆	Ijazah terakhir kepala keluarga	Nominal	1. Tidak punya 2. SD/ setara 3. SLTP/ setara 4. SLTA/ setara 5. Diploma I/ II 6. Akademi ke atas
X ₁₇	Penghasilan rata-rata tiap bulan (Rp)	Rasio	-
X ₁₈	Kepemilikan aset (Rp)	Rasio	-

Berikut ini adalah penjelasan dari masing-masing variabel tersebut.

a. Status kemiskinan rumah tangga miskin (Y)

Rumah tangga adalah seseorang atau sekelompok orang yang mendiami sebagian atau seluruh bangunan fisik atau sensus, dan biasanya tinggal bersama serta makan dari satu dapur. Rumah tangga umumnya terdiri dari ibu, bapak, anak, orang

tua/mertua, famili, pembantu dan lainnya. BPS menetapkan kategori status kemiskinan rumah tangga miskin sebagai berikut (BPS PPLS, 2011).

- i. Golongan Rumah Tangga Sangat Miskin (RTSM) adalah rumah tangga yang mengkonsumsi makanan senilai sampai dengan 1.900 kalori per hari, yang senilai dengan Rp 120.000,- per minggu atau bila disetarakan dengan pengeluaran per bulan adalah Rp 480.000,- per rumah tangga per bulan.
 - ii. Golongan Rumah Tangga Miskin (RTM) adalah rumah tangga yang mengkonsumsi makanan senilai sampai 2.100 kalori per hari, yang senilai dengan Rp 150.000,- per minggu atau bila disetarakan dengan pengeluaran per bulan adalah Rp 600.000,- per rumah tangga per bulan.
 - iii. Golongan Rumah Tangga Hampir Miskin (RTHM) yaitu rumah tangga yang mengkonsumsi makanan senilai sampai dengan 2.300 kalori per hari, yang senilai sampai dengan Rp 175.000,- per minggu atau bila disetarakan dengan pengeluaran per bulannya adalah Rp 700.000,- per rumah tangga per bulan.
- b. Status penguasaan bangunan tempat tinggal (X_1)
Kepemilikan bangunan tempat tinggal dapat mencerminkan kemampuan ekonomi suatu rumah tangga. Bangunan yang ditempati rumah tangga bisa dikategorikan dalam status milik sendiri, kontrak, sewa, bebas sewa, rumah dinas, rumah milik orang tua dan atau sanak saudara, serta selain daripada itu, misalnya rumah adat atau tempat tinggal bersama (BPS PPLS, 2011).
- c. Luas kavling termasuk bangunan dalam satuan m^2 (X_2)
Kavling adalah bagian tanah yang sudah dipetak-petak dengan ukuran tertentu yang akan dijadikan bangunan atau tempat tinggal. Pada umumnya rumah tangga yang termasuk miskin, kondisi fisik atau fasilitas rumah kurang memenuhi. Suatu rumah dapat dikatakan memenuhi salah satu persyaratan sehat jika luas kavling minimal $60 m^2$ untuk setiap rumah tangga ideal terdiri dari 4 orang (BPS PPLS, 2011).

- d. Luas lantai dalam satuan m^2 (X_3)
 Luas Lantai adalah jumlah luas lantai dari setiap bagian bangunan (sebatas atap) yang ditempati (dihuni) dan digunakan untuk keperluan sehari-hari oleh rumah tangga, termasuk teras, garasi, tempat mencuci, WC, gudang, lantai setiap tingkat untuk bangunan bertingkat. Luas lantai tempat tinggal rumah tangga tidak termasuk ruangan khusus untuk usaha, warung, restoran, toko, salon, kandang ternak, lantai jemur, lumbung padi dan lain-lain. Salah satu kriteria rumah tangga miskin yaitu luas lantai rumah kecil, dimana Departemen Kesehatan menentukan bahwa suatu rumah dapat dikatakan memenuhi salah satu persyaratan sehat jika luas lantai rumah minimal $8 m^2$ per orang. Setiap rumah tangga ideal terdiri dari 4 orang, yaitu orang tua dan 2 orang anak, sehingga setiap rumah tangga minimal memiliki luas lantai $32 m^2$ (BPS PPLS, 2011).
- e. Jenis atap rumah terluas (X_4)
 Atap adalah penutup bagian atas suatu bangunan sehingga orang yang mendiami di bawahnya terlindung dari terik matahari, hujan dan sebagainya. Untuk bangunan bertingkat, atap yang dimaksud adalah bagian teratas dari bangunan tersebut. Salah satu fasilitas tempat tinggal yang menunjukkan suatu rumah tangga miskin yaitu atap terbuat dari ijuk/rumbia atau genteng atau seng atau asbes dengan kondisi tidak baik atau kualitas rendah (BPS SPDKP, 2007).
- f. Jenis dinding terluas (X_5)
 Dinding adalah sisi luar atau batas dari suatu bangunan atau penyekat dengan bangunan fisik lainnya. Bila bangunan tersebut menggunakan lebih dari satu jenis dinding yang luasnya sama, maka yang dianggap sebagai dinding terluas adalah dinding yang bernilai lebih tinggi. Salah satu fasilitas tempat tinggal yang menunjukkan suatu rumah tangga termasuk miskin yaitu dinding rumahnya terbuat dari bambu atau kayu atau tembok dengan kondisi tidak baik atau kualitas rendah, termasuk tembok yang sudah usang atau berlumut atau tembok tidak diplester (BPS SPDKP, 2007).

- g. Jenis lantai terluas (X_6)
Jenis lantai terdiri dari keramik atau marmer atau granit, ubin/ tegel/ teras, semen/ bata merah, kayu berkualitas rendah, bambu, tanah. dan lainnya. Lantai ubin yang dilapisi karpet atau vinil tetap dikategorikan ubin. Jika lantai bangunan tempat tinggal lebih dari satu jenis, dipilih yang terluas. Salah satu fasilitas tempat tinggal yang menunjukkan suatu rumah tangga termasuk miskin yaitu sebagian besar lantai terbuat dari tanah atau kayu atau semen atau keramik dengan kondisi tidak baik atau kualitas rendah (BPS SPDKP, 2007).
- h. Fasilitas ketersediaan tempat buang air besar (jamban) (X_7)
Ketersediaan tempat buang air besar atau jamban yang dapat digunakan rumah tangga menjadi salah satu indikator yang penting dalam mendukung pola hidup sehat dan mencerminkan taraf kesejahteraan. Disamping ada tidaknya jamban, indikator penggunaan fasilitas jamban juga penting yang dibedakan atas jamban sendiri, jamban bersama, jamban umum dan tidak ada. Salah satu fasilitas tempat tinggal yang menunjukkan suatu rumah tangga termasuk miskin yaitu tidak memiliki fasilitas buang air besar atau bersama-sama dengan rumah tangga lain (BPS PPLS, 2011).
- i. Tempat pembuangan akhir tinja (X_8)
Salah satu fasilitas tempat tinggal yang menunjukkan suatu rumah tangga termasuk miskin yaitu ketersediaan tempat buang akhir tinja yang dapat digunakan. Tempat pembuangan akhir tinja dapat dibedakan atas septictank, kolam/ sawah, sungai/ danau/ laut, lobang tanah dan selain daripada itu (BPS PPLS, 2011).
- j. Sumber penerangan utama (X_9)
Sumber penerangan utama dibedakan menjadi empat, yaitu listrik PLN meteran, listrik PLN tanpa meteran, listrik non PLN, dan bukan listrik. Listrik non PLN mencakup sumber penerangan listrik yang dikelola oleh instansi atau pihak lain selain PLN, termasuk yang menggunakan sumber penerangan dari aki, generator, dan pembangkit listrik tenaga surya yang dikelola bukan oleh PLN. Bukan listrik seperti lampu

gas elpiji (LPG) dan biogas yang dibangkitkan sendiri maupun berkelompok, sumber penerangan dari minyak tanah (petromak/ lampu tekan, aladin, teplok, sentir, pelita, dan sejenisnya) dan lainnya meliputi lampu karbit, lilin, biji jarak dan kemiri. Salah satu fasilitas tempat tinggal yang menunjukkan suatu rumah tangga termasuk miskin yaitu penerangan bangunan tempat tinggal bukan dari listrik atau listrik tanpa meteran (BPS SPDKP, 2007).

k. Sumber air minum (X_{10})

Sumber air minum dapat dibedakan menjadi tujuh yaitu air dalam kemasan, ledeng, pompa, sumur, mata air, air sungai dan selain daripada itu. Salah satu fasilitas tempat tinggal yang menunjukkan suatu rumah tangga termasuk miskin yaitu sumber air minum berasal dari sumur atau mata air tak terlindung atau dari air hujan (BPS SPDKP, 2007).

l. Bahan bakar memasak (X_{11})

Bahan bakar memasak (BBM) yang digunakan rumah tangga merupakan indikator yang dapat menunjukkan kesejahteraan kehidupan rumah tangga tersebut. BBM dapat dibedakan menjadi listrik, gas/ elpiji, minyak tanah, arang kayu/ tempurung, kayu bakar dan lainnya. Salah satu fasilitas tempat tinggal yang mencerminkan suatu rumah tangga termasuk miskin yaitu bahan bakar untuk memasak sehari-hari adalah kayu bakar atau arang atau minyak tanah (BPS PPLS, 2011).

m. Intensitas konsumsi daging/ susu/ ayam per minggu (X_{12})

Pola konsumsi makanan yang diterapkan dalam suatu rumah tangga dapat mencerminkan kemampuan ekonomi ataupun kesejahteraan rumah tangga tersebut. Salah satu kriteria umum rumah tangga yang termasuk miskin yaitu sebagian besar pengeluarannya digunakan untuk memenuhi konsumsi makanan pokok yang sangat sederhana serta umumnya hanya mengkonsumsi daging/ susu/ ayam satu kali dalam seminggu (BPS SPDKP, 2007).

- n. Intensitas membeli pakaian per tahun (X_{13})
Salah satu kriteria umum rumah tangga yang termasuk miskin yaitu hanya membeli satu stel pakaian baru dalam setahun (BPS SPDKP, 2007).
- o. Intensitas makan per hari (X_{14})
Salah satu kriteria umum rumah tangga yang termasuk miskin yaitu hanya sanggup makan sebanyak satu atau dua kali dalam sehari (BPS SPDKP, 2007).
- p. Pengobatan (X_{15})
Salah satu kriteria umum rumah tangga yang termasuk miskin yaitu biasanya tidak mampu atau mengalami kesulitan untuk berobat ke tenaga medis kecuali puskesmas atau yang disubsidi pemerintah (BPS SPDKP, 2007).
- q. Ijazah terakhir kepala keluarga (X_{16})
Semakin tinggi pendidikan terakhir yang ditamatkan, mengindikasikan baiknya kualitas pendidikan dan kemampuan orang tersebut. Salah satu kriteria umum rumah tangga yang termasuk miskin yaitu pendidikan tertinggi kepala rumah tangga tidak sekolah atau tidak tamat SD atau hanya SD (BPS PPLS, 2011).
- r. Penghasilan rata-rata per bulan (X_{17})
Salah satu kriteria orang disebut miskin adalah jika sumber penghasilan kepala rumah tangga miskin adalah: petani dengan luas lahan 0,5 ha, buruh tani, nelayan, buruh bangunan, buruh perkebunan, atau pekerjaan lainnya dengan pendapatan < Rp 600.000 per bulan (BPS PPLS, 2011).
- s. Kepemilikan aset di dalam rumah dalam satuan Rupiah (X_{18})
Kepemilikan aset dalam rumah tempat tinggal sendiri dapat mencerminkan kemampuan ekonomi suatu rumah tangga. Salah satu kriteria orang disebut miskin adalah jika tidak memiliki tabungan atau barang yang mudah dijual dengan minimal Rp 500.000,- seperti emas, ternak, motor, atau barang modal lainnya (BPS PPLS, 2011).

3.3 Metode Analisis

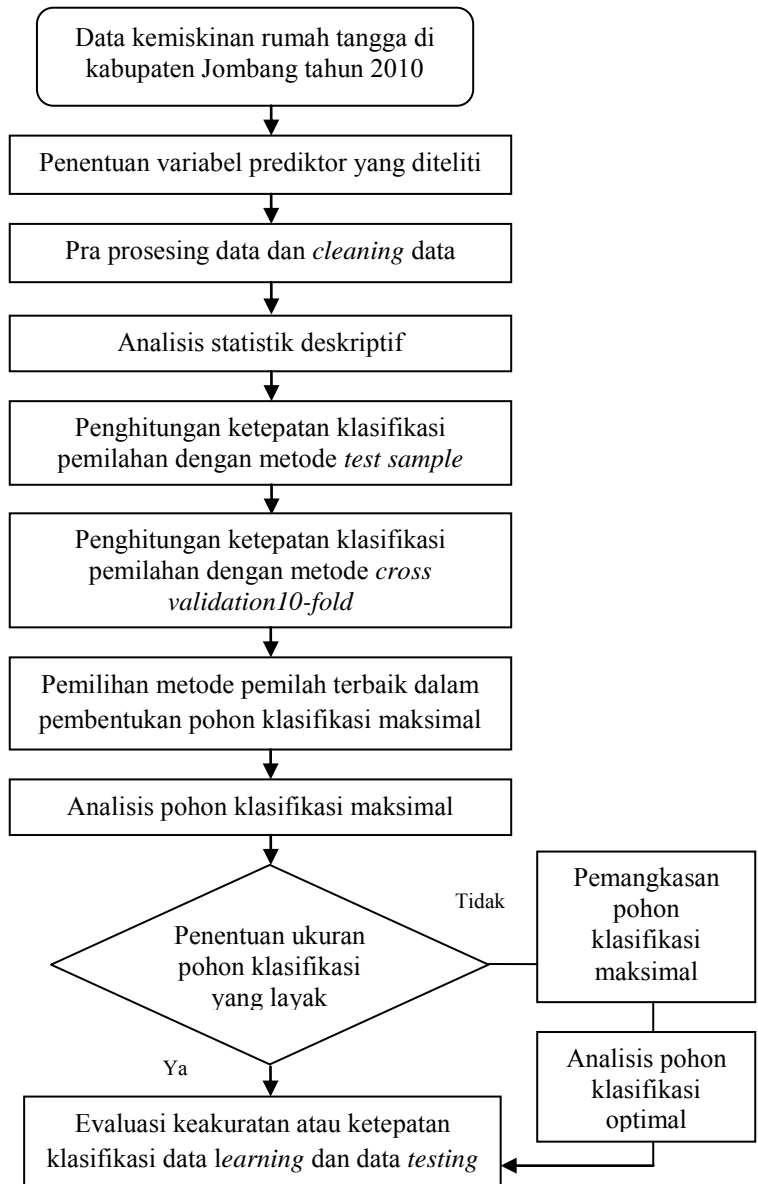
Metode yang digunakan dalam analisis data untuk menjawab permasalahan penelitian ini adalah pendekatan CART.

Berikut ini tahapan yang dilakukan dalam menganalisis data penelitian.

- 1) Mendapatkan data keseluruhan hasil survei verifikasi rumah tangga miskin dari Bappeda Kabupaten Jombang.
- 2) Memilih variabel yang akan digunakan dalam penelitian, sebagaimana diuraikan pada subbab 3.2.
- 3) Pra prosesing data yang sudah terkumpul (73.720 data) dengan melakukan pengkodean data pada setiap variabel bertipe kategorik sesuai dengan pengkategorian yang telah ditetapkan. Selain itu juga melakukan *cleaning* data terhadap data-data pengamatan yang banyak ditemukan kosong atau tidak diisi secara lengkap oleh responden sehingga menyebabkan banyak informasi yang kurang dari unit pengamatan tersebut. Lalu menyusun data yang siap diolah sesuai struktur data CART.
- 4) Menyajikan statistik deskriptif data untuk memberikan gambaran awal tentang karakteristik data penelitian dan variabel yang diteliti. Jika variabel prediktor berskala rasio, dihitung nilai rata-rata, standar deviasi, nilai minimum dan nilai maksimum. Sementara jika variabel prediktor berskala nominal atau ordinal dibuat *pie chart* per variabel menurut setiap kelas variabel respon. Selain itu juga dibuat *bar chart* tentang perbandingan jumlah masing-masing kelas variabel respon (RTSM, RTM dan RTHM) per kecamatan di kabupaten Jombang.
- 5) Membagi sejumlah data yang sudah fix hasil dari *cleaning* dan pra-prosesing data (43.544 data) menjadi data *learning* dan data *testing* dengan kombinasi proporsi tertentu (mengacu dari penelitian-penelitian sebelumnya) yaitu 95 persen : 5 persen, 90 persen : 10 persen, 85 persen : 15 persen, 80 persen : 20 persen, 75 persen : 25 persen dan 70 persen : 30 persen. Masing-masing kombinasi proporsi tersebut diolah untuk mencobakan alternatif metode pemilahan *test sample estimate*. Sehingga diperoleh suatu nilai ketepatan klasifikasi dan banyak simpul terminal yang terbentuk dari masing-masing kombinasi proporsi data *learning-testing* tersebut.

- 6) Mencoba alternatif lain metode pemilahan yaitu *cross validation V-fold estimate*, dimana dalam penelitian ini nilai *V* yang digunakan adalah 10. Sejumlah data yang sudah fix hasil dari *cleaning* dan pra prosesing data (43.544 data) diolah dengan metode pemilahan *cross validation 10-fold* dengan indeks Gini dan indeks Twoing. Sehingga diperoleh suatu nilai ketepatan klasifikasi dan banyak simpul terminal yang terbentuk.
- 7) Membandingkan nilai ketepatan klasifikasi data *testing* hasil langkah 5) dan 6). Metode pemilahan yang menghasilkan nilai ketepatan klasifikasi data *testing* terbesar dengan jumlah simpul terminal relatif sederhana adalah yang nantinya dipilih dalam pembuatan pohon klasifikasi maksimal.
- 8) Melakukan analisis pembentukan pohon klasifikasi maksimal.
- 9) Menentukan ukuran pohon klasifikasi yang layak dengan melihat besarnya nilai kompleksitas pohon klasifikasi yang terbentuk dan nilai *resubstitution relative cost*. Jika nilai kompleksitas 0,000 dan nilai *resubstitution relative cost* yang kecil (menunjukkan struktur data dari pohon klasifikasi maksimal kompleks) maka perlu dilakukan pemangkasan pohon klasifikasi maksimal (*pruning*) sehingga diperoleh suatu pohon klasifikasi optimal.
- 10) Melakukan analisis pohon klasifikasi optimal yang terbentuk.
- 11) Mendapatkan karakteristik kelas simpul terminal-simpul terminal yang dihasilkan dari penelusuran pohon klasifikasi optimal.
- 12) Menghitung nilai 1-APER yang dihasilkan oleh data *learning* dan data *testing* dari pohon klasifikasi optimal untuk melihat kebaikan dan keakuratan pohon klasifikasi optimal tersebut.

Tahapan analisis data di atas disajikan dalam bentuk *flow chart* sebagai berikut.



Gambar 3.1 Flow chart langkah analisis.

(Halaman ini sengaja dikosongkan)

BAB IV

ANALISIS DATA DAN PEMBAHASAN

Pada bab ini dilakukan ulasan hasil pengolahan data dan analisis data untuk menjawab permasalahan dalam penelitian. Pada subbab 4.1 diberikan penjelasan hasil statistik deskriptif data penelitian yaitu tentang karakteristik rumah tangga miskin yang ada di Kabupaten Jombang tahun 2010 untuk memberikan gambaran mengenai unit analisis yang diteliti. Lalu pada subbab 4.2 diberikan penjelasan analisis klasifikasi rumah tangga miskin dengan pendekatan pohon klasifikasi CART.

4.1 Karakteristik Rumah Tangga Miskin di Kabupaten Jombang Tahun 2010

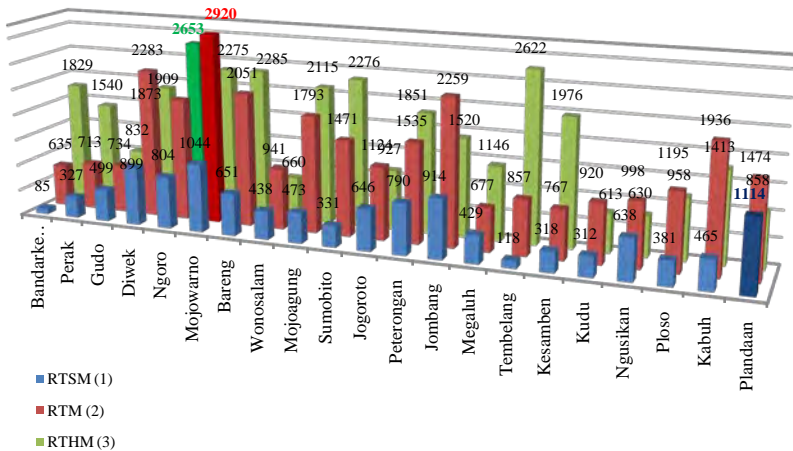
Sebagaimana dijelaskan pada subbab 3.2 bahwa BPS membagi kategori status kemiskinan rumah tangga miskin menjadi tiga berdasarkan pengeluaran rata-rata per bulan yaitu kategori Rumah Tangga Sangat Miskin (RTSM), Rumah Tangga Miskin (RTM) dan Rumah Tangga Hampir Miskin (RTHM). Dari hasil survei verifikasi rumah tangga miskin pada tahun 2010 terhadap 73.720 rumah tangga miskin yang tersebar di 21 kecamatan di Kabupaten Jombang, diperoleh informasi berupa data status kemiskinan masing-masing rumah tangga miskin dan juga data-data terkait delapan belas variabel lainnya yang diteliti. Selanjutnya data tentang status kemiskinan rumah tangga miskin yang telah diperoleh tersebut dikoding menjadi tiga kelas yaitu RTSM (1), RTM (2) dan RTHM (3) untuk dijadikan sebagai variabel respon dalam penelitian ini. Dari data tersebut, diperoleh informasi bahwa banyaknya setiap kelas rumah tangga miskin di masing-masing kecamatan tidak sama. Data secara rinci diberikan dalam Tabel 4.1 dan digambarkan dalam bentuk diagram batang 3 dimensi pada Gambar 4.1. Diantara 73.720 rumah tangga miskin yang ada, 11.676 rumah tangga atau sekitar 15,8 persen termasuk kelas RTSM, 29.156 rumah tangga (39,6 persen) termasuk RTM, dan paling banyak yaitu 32.888 rumah tangga (44,6 persen) termasuk kelas RTHM.

Tabel 4.1 Data Banyaknya Rumah Tangga Miskin pada Setiap Kelas per Kecamatan di Kabupaten Jombang Tahun 2010

No.	Kecamatan	Jumlah RTSM	Jumlah RTM	Jumlah RTHM	Total
1	Bandarkedungmulyo	85	635	1.829	2.549
2	Perak	327	713	1.540	2.580
3	Gudo	499	734	832	2.065
4	Diwek	899	2.283	1.909	5.091
5	Ngoro	804	1.873	2.653 ^{*)}	5.330
6	Mojowarno	1.044	2.920 ^{*)}	2.275	6.239
7	Bareng	651	2.051	2.285	4.987
8	Wonosalam	438	941	660	2.039
9	Mojoagung	473	1.793	2.115	4.381
10	Sumobito	331	1.471	2.276	4.078
11	Jogoroto	646	1.124	927	2.697
12	Peterongan	790	1.535	1.851	4.176
13	Jombang	914	2.259	1.520	4.693
14	Megaluh	429	677	1.146	2.252
15	Tembelang	118	857	2.622	3.597
16	Kesamben	318	767	1.976	3.061
17	Kudu	312	920	613	1.845
18	Ngusikan	638	998	630	2.266
19	Ploso	381	1.195	958	2.534
20	Kabuh	465	1.936	1.413	3.814
21	Plandaan	1.114 ^{*)}	1.474	858	3.446
Total		11.676	29.156	32.888	73.720

^{*)} Jumlah terbanyak suatu kelas rumah tangga miskin

Data yang diberikan dalam Tabel 4.1 digambarkan dalam Gambar 4.1 agar informasi yang disampaikan dapat dilihat secara visual dengan lebih jelas dan menarik. Pada setiap kecamatan dapat diketahui perbandingan banyaknya rumah tangga miskin yang termasuk ke dalam kelas RTSM, RTM maupun RTHM.



Gambar 4.1 Banyaknya rumah tangga miskin pada setiap kelas per kecamatan di Kabupaten Jombang.

Gambar 4.1 memberikan informasi bahwa diantara dua puluh satu kecamatan di kabupaten Jombang, ada sebelas kecamatan yang memiliki jumlah rumah tangga miskin paling banyak termasuk ke dalam kelas RTHM dibandingkan kelas RTM dan RTSM. Hal ini ditunjukkan dengan gambar diagram hijau yang tingginya melebihi diagram batang merah dan biru dalam satu kecamatan. Sebelas kecamatan tersebut meliputi Kecamatan Bandarkedungmulyo, Perak, Gudo, Ngoro, Bareng, Mojowagung, Sumobito, Peterongan, Megaluh, Tembelang, dan Kesamben. Sementara sepuluh kecamatan lainnya memiliki jumlah rumah tangga miskin paling banyak termasuk ke dalam RTM dibandingkan kelas RTHM dan RTSM, ditunjukkan dengan gambar diagram merah yang tingginya melebihi diagram batang hijau dan biru dalam satu kecamatan. Sepuluh kecamatan tersebut meliputi Kecamatan Diwek, Mojowarno, Wonosalam, Jogoroto, Jombang, Kudu, Ngusikan, Ploso, Kabuh, dan Plandaan. Rumah tangga miskin dengan kelas RTSM paling banyak ada di kecamatan Plandaan (biru tua), kelas RTM paling banyak ada di kecamatan Mojowarno (merah tua), dan kelas RTHM paling banyak ada di kecamatan Ngoro (hijau tua).

Secara keseluruhan, data yang diperoleh dari survei verifikasi rumah tangga miskin di Kabupaten Jombang tahun 2010 pada kenyataannya menunjukkan bahwa tidak semua unit analisis yang diteliti mampu memberikan data-data atau informasi secara jelas, tepat, dan lengkap sebagaimana yang diharapkan. Sebab, pada kedelapan belas variabel prediktor yang diteliti, banyak diantaranya ditemukan data-data *missing* maupun *outlier*. Maka dari itu, sebelum dilakukan analisis klasifikasi terlebih dahulu dilakukan pra pemrosesan data. Pra pemrosesan data meliputi *filtering* dan *sorting* data pada masing-masing variabel dan unit analisis, dengan ketentuan yaitu data-data yang mengandung informasi tidak lengkap tidak diikuti sertakan dalam analisis berikutnya. Diperoleh banyaknya data pengamatan yang siap dianalisis lebih lanjut ada sebanyak 43.544 unit rumah tangga miskin. Selesai dilakukan pra pemrosesan data, selanjutnya terhadap data yang siap tersebut dibuat analisis statistik deskriptif dengan cara menghitung nilai rata-rata, standar deviasi, nilai minimum dan nilai maksimum untuk variabel prediktor yang berskala rasio, serta membuat *pie chart* untuk variabel prediktor yang berskala nominal. Berikut ini diberikan ulasan hasil statistik deskriptif data.

Tabel 4.2 Statistik Deskriptif Variabel Prediktor Berskala Rasio

Variabel		Rata-rata	St. Dev	Min	Maks
Luas Kavling (m ²)	X ₂	112,31	107,31	0	600
Luas Lantai (m ²)	X ₃	43,642	28,133	0	320
Konsumsi Daging Susu Ayam (kali per minggu)	X ₁₂	0,738	0,743	0	7
Konsumsi Satu set Pakaian (kali per tahun)	X ₁₃	0,823	0,478	0	4
Konsumsi Makan (kali per hari)	X ₁₄	2,917	0,287	1	3
Penghasilan rata-rata per bulan (Rp)	X ₁₇	238.798	244.351	0	3.000.000
Kepemilikan Aset (Rp)	X ₁₈	611.262	1.593.001	0	15.000.000

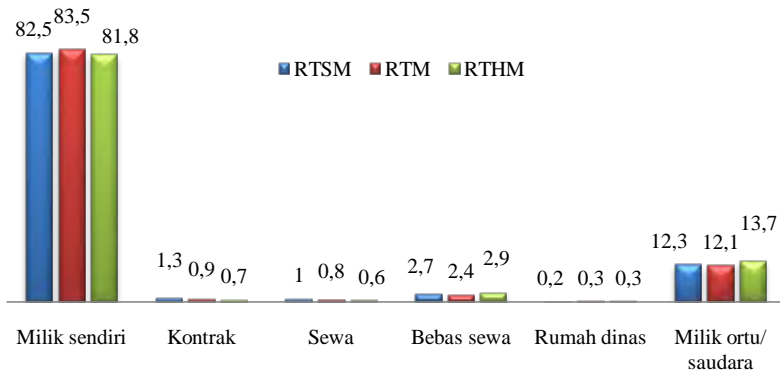
Berdasarkan Tabel 4.2 dapat diperoleh beberapa informasi antara lain yaitu luas kavling bangunan tempat tinggal rumah tangga miskin di Kabupaten Jombang rata-rata sebesar $112,31 \text{ m}^2$. Standar deviasi menunjukkan sebaran data luas kavling rumah tangga miskin di Kabupaten Jombang cukup tinggi yaitu 107,31. Luas kavling minimal 0 m^2 dan maksimal 600 m^2 . Rumah tangga miskin dengan luas kavling minimal 0 m^2 tersebut yaitu rumah tangga miskin dari kelas RTSM maupun RTM yang tidak mempunyai kavling bangunan tempat tinggal. Kemudian luas lantai tempat tinggal rumah tangga miskin di Kabupaten Jombang, rata-rata mempunyai luas $43,642 \text{ m}^2$ dengan sebaran data luas lantai rumah tangga miskin disana yaitu 28,133. Luas lantai minimal yaitu 0 m^2 dan maksimal 320 m^2 . Sama halnya dengan luas kavling, luas lantai minimal 0 m^2 tersebut yaitu rumah tangga miskin dari kelas RTSM, RTM maupun RTHM yang tidak mempunyai lantai bangunan tempat tinggal. Dalam hal ini yaitu rumah tangga miskin yang tidak mempunyai tempat tinggal sendiri, sehingga dianggap bahwa rumah tangga tersebut tidak memiliki lantai bangunan tempat tinggal.

Intensitas konsumsi daging susu ayam per minggu oleh suatu rumah tangga miskin di Kabupaten Jombang rata-rata hanya $0,738 \approx 1$ kali dalam seminggu dengan ukuran sebaran data tersebut relatif kecil yaitu 0,743. Dalam seminggu, ada rumah tangga miskin dari kelas RTSM yang tidak mengkonsumsi daging susu ayam. Maksimal suatu rumah tangga miskin mengkonsumsi daging susu ayam sebanyak 7 kali dalam seminggu. Sementara untuk intensitas konsumsi satu set pakaian per tahun, rata-rata suatu rumah tangga miskin di Kabupaten Jombang membeli pakaian satu set yaitu $0,823 \approx 1$ kali dalam setahun dengan ukuran sebaran data yang ada juga relatif kecil yaitu 0,478. Hal ini menunjukkan bahwa keragaman data intensitas konsumsi pakaian oleh rumah tangga miskin yang termasuk kelas mana pun, yang ada di Kabupaten Jombang relatif kecil variasinya. Suatu rumah tangga miskin paling banyak mengkonsumsi satu set pakaian 4 kali dalam setahun. Untuk intensitas konsumsi makan per hari, rata-rata suatu rumah tangga miskin di Kabupaten Jombang makan $2,917 \approx 3$ kali dalam sehari dengan standar deviasi data

kecil yaitu 0,287. Minimal suatu rumah tangga miskin makan sekali dalam sehari dan maksimal makan 3 kali sehari.

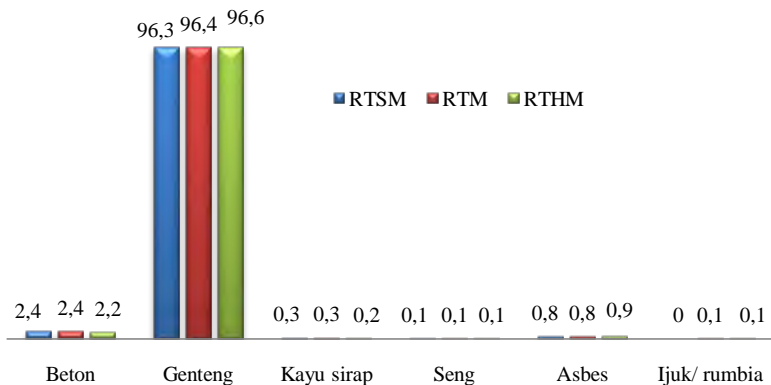
Penghasilan rata-rata per bulan rumah tangga miskin di kabupaten Jombang sebesar Rp 238.798,- dengan standar deviasi data relatif tinggi yaitu 244.351, yang menunjukkan bahwa data penghasilan rata-rata per bulan rumah tangga miskin disana cukup banyak variasi nilainya. Penghasilan rata-rata per bulan minimal Rp 0,- dan maksimal Rp 3.000.000,-. Rumah tangga miskin dengan penghasilan Rp 0,- tersebut besar kemungkinan merupakan rumah tangga miskin dengan kelas RTSM yang sehari-hari tidak memiliki pekerjaan tetap. Sementara itu, untuk kepemilikan asset yang dihitung dalam satuan rupiah, rata-rata asset yang dimiliki rumah tangga miskin di Kabupaten Jombang sebesar Rp 611.262,-. Standar deviasi data sangat tinggi yaitu 1.593.001 yang berarti bahwa data kepemilikan asset rumah tangga miskin disana sangat bervariasi nilainya, dengan nilai minimal Rp 0,- dan maksimal asset yang dimiliki senilai Rp 15.000.000,-. Rumah tangga miskin yang memiliki asset Rp 0,- atau dapat dikatakan tidak memiliki asset besar kemungkinan merupakan rumah tangga miskin yang memang sehari-hari kehidupannya tidak menentu, tempat tinggalnya menumpang atau bukan milik sendiri, penghasilannya juga tidak menentu sehingga dianggap rumah tangga miskin tersebut tidak memiliki asset.

Karakteristik lain dari rumah tangga miskin di Kabupaten Jombang ditinjau pada variabel-variabel prediktor yang berskala nominal sebagaimana disebutkan dalam subbab 3.2, digambarkan dalam *pie chart* per masing-masing kelas rumah tangga miskin. Ada sebelas *pie chart* yang ditampilkan, meliputi status kepemilikan bangunan, jenis atap terluas, jenis dinding terluas, jenis lantai terluas, fasilitas tempat buang air besar (jamban), tempat pembuangan akhir tinja, sumber penerangan utama, sumber air minum, bahan bakar minyak, pengobatan, dan ijazah terakhir kepala keluarga pada masing-masing kelas rumah tangga miskin yaitu RTSM, RTM, dan RTHM. Berikut ini diberikan penjelasan dari masing-masing variabel tersebut.



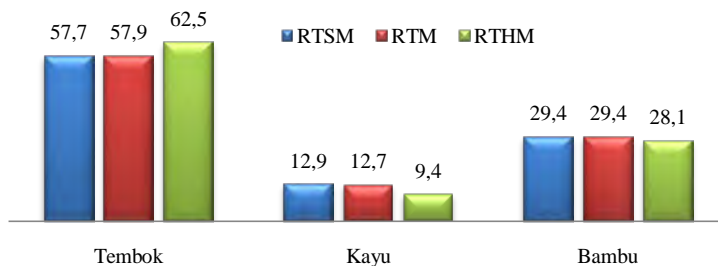
Gambar 4.2 Prosentase rumah tangga miskin per kelas menurut status kepemilikan bangunan.

Mengenai status kepemilikan bangunan, dari Gambar 4.2 dapat dilihat bahwa pada semua kelas rumah tangga miskin menunjukkan mayoritas bangunan yang dihuni atau ditempati adalah milik sendiri (kategori 1) dengan persentase mencapai 82,5 persen pada RTSM, 83,5 persen pada RTM, dan 81,8 persen pada RTHM. Persentase tertinggi berikutnya menunjukkan bahwa bangunan yang ditempati adalah milik orang tua/ sanak saudara (kategori 6). Status kepemilikan selain kategori 1 dan 6 memiliki persentase relatif sangat kecil.



Gambar 4.3 Prosentase rumah tangga miskin per kelas menurut jenis atap terluas.

Jenis atap terluas dari bangunan tempat tinggal ketiga kelas rumah tangga miskin menunjukkan bahwa mayoritas bangunan tempat tinggal adalah beratap jenis genteng (kategori 2) dengan persentase mencapai 96,3 persen pada RTSM, 96,4 persen pada RTM, dan 96,6 persen pada RTHM. Sedangkan untuk lima kategori lainnya memiliki persentase yang terpaut sangat jauh dan nilainya sangat kecil sebagaimana ditampilkan oleh Gambar 4.3.

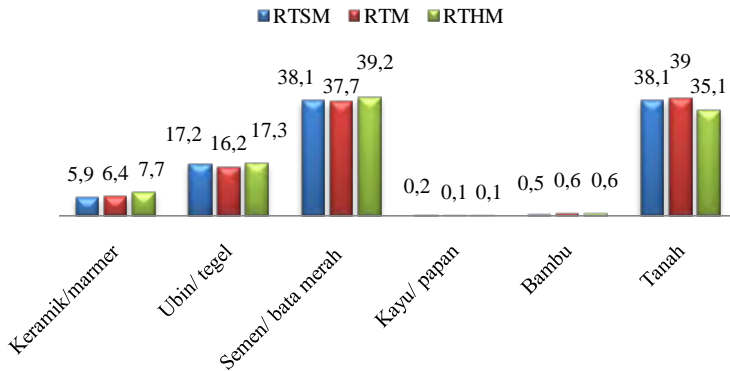


Gambar 4.4 Prosentase rumah tangga miskin per kelas menurut jenis dinding terluas.

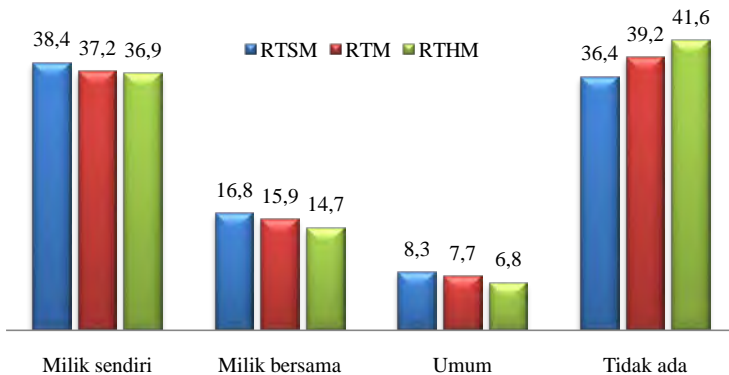
Mengenai jenis dinding terluas, pada semua kelas rumah tangga miskin menunjukkan bahwa mayoritas bangunan yang ditempati memiliki dinding berupa tembok (kategori 1) dengan persentase mencapai 57,7 persen pada RTSM, 57,9 persen pada RTM, dan 62,5 persen pada RTHM. Lalu persentase tertinggi berikutnya yaitu bangunan berdinding bambu, dimana ada 28,1 persen untuk RTHM, sedangkan RTM dan RTSM masing-masing sekitar 29,4 persen. Jenis dinding kayu memiliki nilai persentase paling kecil pada ketiga kelas (Gambar 4.4).

Pada Gambar 4.5, dapat dilihat bahwa jenis lantai terluas dari bangunan hunian atau tempat tinggal untuk kelas RTHM paling banyak adalah dari semen/ bata merah (kategori 3) dengan persentase 39,2 persen dan persentase terbanyak berikutnya berlantai tanah (kategori 6) yaitu ada 35,1 persen. Kebalikan dari RTHM, untuk kelas RTM lebih banyak yang memiliki tempat tinggal berlantai tanah daripada berlantai semen/ bata merah, masing-masing 39 persen dan 37,7 persen. Sementara untuk kelas

RTSM, persentase yang berlantai tanah sama besarnya dengan yang berlantai semen/ bata merah yaitu sekitar 38,1 persen. Pada semua kelas, sedikit yang memiliki bangunan berlantai ubin/tegel (kategori 2) dan keramik/ marmer (kategori 1) dan hanya sekitar 0.1 persen yang berlantai kayu/ bambu (kategori 4).



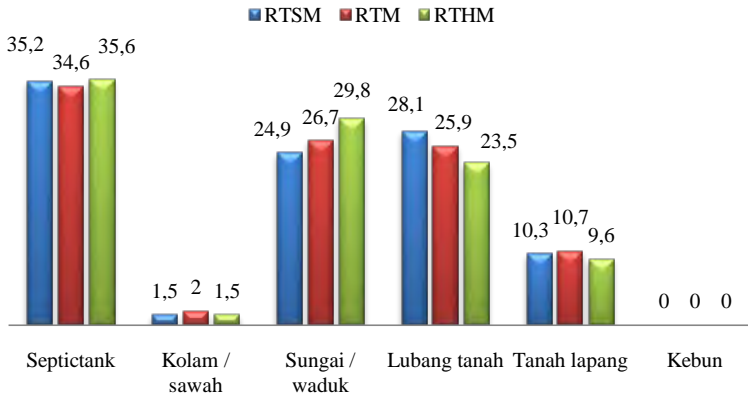
Gambar 4.5 Prosentase rumah tangga miskin per kelas menurut jenis lantai terluas.



Gambar 4.6 Prosentase rumah tangga miskin per kelas menurut fasilitas tempat buang air besar (jamban).

Fasilitas tempat buang air besar (jamban), pada rumah tangga miskin kelas RTHM dan RTM, berturut-turut yang memiliki persentase terbesar adalah kategori 4 (tidak ada),

kategori 1 (milik sendiri), kategori 2 (milik bersama), dan yang terkecil adalah kategori 3 (umum). Namun pada kelas RTSM, persentase jamban milik sendiri lebih tinggi daripada tidak ada fasilitas yaitu masing-masing sebesar 38,4 persen dan 36,4 persen (Gambar 4.6).

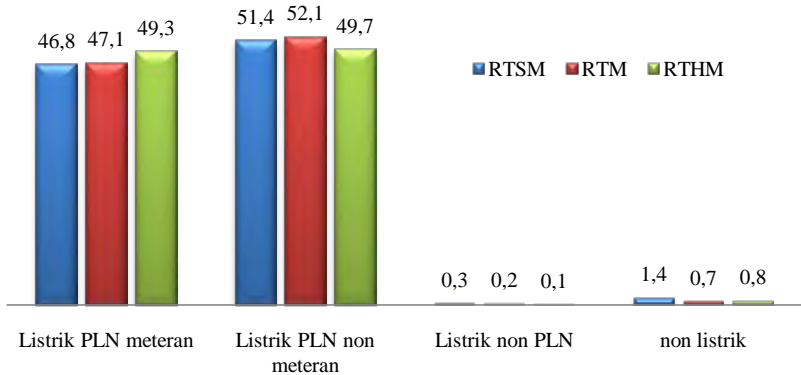


Gambar 4.7 Prosentase rumah tangga miskin per kelas menurut tempat pembuangan akhir tinja.

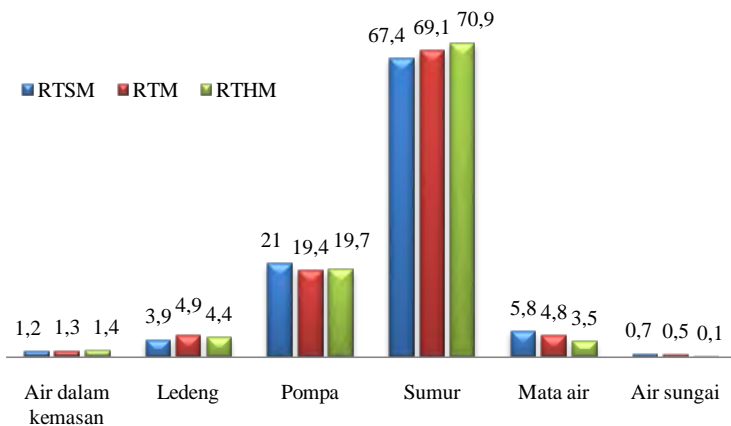
Gambar 4.7 memberikan informasi bahwa tempat pembuangan akhir tinja pada semua kelas rumah tangga miskin di kabupaten Jombang yang memiliki persentase paling besar adalah kategori 1 (septictank). Persentase RTHM yang menggunakan septictank sebanyak 35,6 persen, RTM sebanyak 34,6 persen dan pada RTSM sebanyak 35,2 persen. Tempat pembuangan akhir tinja dengan persentase terbesar kedua adalah kategori 3 (sungai/waduk) dan yang ketiga adalah kategori 4 (lubang tanah) untuk kelas RTHM dan kelas RTM. Namun pada kelas RTSM menunjukkan hasil yang berkebalikan yaitu lebih banyak yang membuang air tinja di lubang tanah dibandingkan di sungai/waduk.

Sumber penerangan utama yang paling banyak digunakan pada semua kelas rumah tangga miskin di Kabupaten Jombang yaitu listrik PLN non meteran dengan persentase sebesar 51,4 persen pada RTSM, 52,1 persen pada RTM dan 49,7 persen pada RTHM. Lalu disusul oleh listrik PLN meteran dengan perbedaan

selisih persentase yang relatif kecil yaitu ada 46,8 persen pada RTSM, 47,1 persen pada RTM, dan 49,3 persen pada RTHM. Sedikit sekali persentase rumah tangga miskin yang masih menggunakan sumber penerangan non listrik dan listrik non PLN (Gambar 4.8).



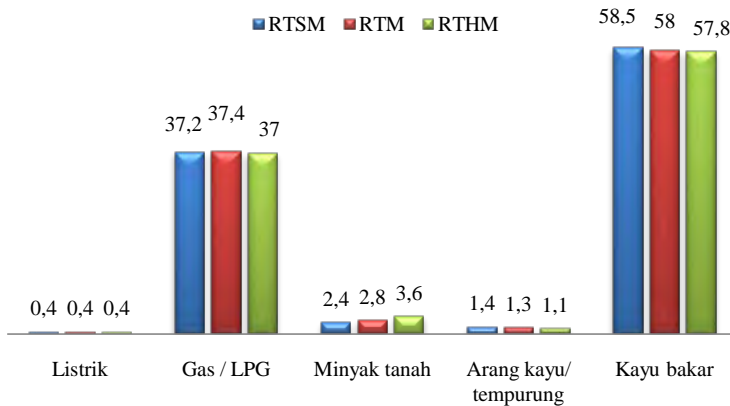
Gambar 4.8 Prosentase rumah tangga miskin per kelas menurut sumber penerangan utama.



Gambar 4.9 Prosentase rumah tangga miskin per kelas menurut sumber air minum.

Gambar 4.9 memberikan informasi tentang sumber air minum yang banyak digunakan oleh RTSM, RTM, maupun

RTHM di kabupaten Jombang yaitu sumur (kategori 4) dengan persentase sebesar 67,4 persen pada kelas RTSM, 69,1 persen pada RTM dan 70,9 persen pada RTHM. Berikutnya yang masih banyak digunakan yaitu pompa dengan persentase sebesar 21 persen pada kelas RTSM, 19,4 persen pada RTM dan 19,7 persen pada RTHM. Sedangkan empat kategori sumber air munum lainnya memiliki persentase relatif kecil dibandingkan sumur dan pompa.

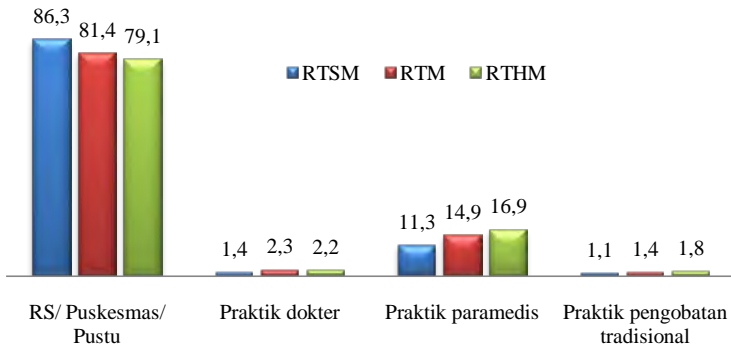


Gambar 4.10 Prosentase rumah tangga miskin per kelas menurut bahan bakar minyak.

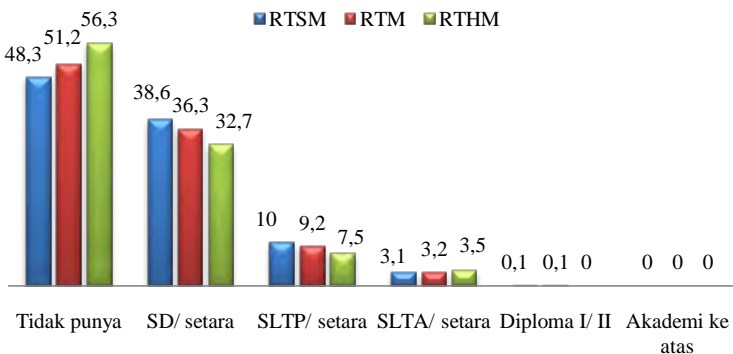
Jenis bahan bakar yang digunakan, baik kelas RTSM, RTM maupun RTHM paling banyak menggunakan kayu bakar (kategori 5) dengan persentase masing-masing 58,5 persen, 58 persen, dan 57,8 persen. Bahan bakar minyak dengan persentase tinggi berikutnya adalah gas/ LPG (kategori 2) yaitu 37,2 persen pada RTSM, 37,4 persen pada RTM dan 37 persen pada RTHM. Sedangkan yang menggunakan listrik, minyak tanah dan arang kayu/ tempurung memiliki persentase cukup rendah (Gambar 4.10).

Ketika membutuhkan suatu pengobatan baik karena sakit atau keperluan medis lainnya, rumah tangga miskin pada semua kelas memilih berobat ke RS/ Puskesmas/Pustu (kategori 1) dengan persentase 86,3 persen pada RTSM, 81,4 persen pada RTM, dan 79,1 persen pada RTHM. Alternatif pengobatan

berikutnya yang memiliki persentase tertinggi digunakan rumah tangga miskin disana adalah pengobatan dari praktik paramedis (kategori 3) dengan persentase 11,3 persen pada RTSM, 14,9 persen pada RTM, dan 16,9 persen pada RTHM. Sementara dua kategori lainnya memiliki persentase yang relatif kecil dibandingkan kategori 1 dan 3 sebagaimana ditampilkan Gambar 4.11 berikut.



Gambar 4.11 Prosentase rumah tangga miskin per kelas menurut pengobatan.



Gambar 4.12 Prosentase rumah tangga miskin per kelas menurut ijazah terakhir kepala keluarga.

Kepala keluarga (KK) rumah tangga miskin di kabupaten Jombang pada semua kelas menunjukkan bahwa mayoritas tidak mempunyai ijazah pendidikan terakhir (kategori 1). Pada RTSM, persentase KK yang tidak mempunyai ijazah pendidikan terakhir

sebesar 48,3 persen, pada RTM sebesar 51,2 persen dan pada RTHM sebesar 56,3 persen. Berikutnya ada sekitar 38,6 persen RTSM yang memiliki ijazah terakhir SD/ setara (kategori 2), pada RTM ada 36,3 persen dan pada RTHM ada sebanyak 32,7 persen (Gambar 4.12). Untuk taraf pendidikan yang lebih tinggi yaitu SLTP/ setara, SLTA/setara, Diploma, Akademi ke atas justru memiliki persentase yang kecil. Hal ini cukup memprihatinkan karena ada indikasi tingkat pendidikan rumah tangga miskin yang menjadi unit analisis dalam penelitian ini masih sangat rendah.

4.2 Klasifikasi Rumah Tangga Miskin di Kabupaten Jombang dengan Pohon Klasifikasi CART

Mengacu pada tujuan penelitian ini yaitu ingin diperoleh informasi mengenai variabel prediktor apa yang menjadi faktor paling penting dalam menentukan klasifikasi status kemiskinan rumah tangga miskin di kabupaten Jombang, maka berikut ini dilakukan analisis klasifikasi status kemiskinan rumah tangga miskin di kabupaten Jombang menggunakan pendekatan CART. Variabel respon berupa data kategorik yaitu status kemiskinan rumah tangga miskin dengan tiga level kelas (RTSM (1), RTM (2) dan RTHM (3)) sehingga metode analisis pendekatan CART akan menghasilkan suatu pohon klasifikasi.

Sesuai dengan prosedur algoritma CART, tahapan pertama yang dilakukan adalah pembentukan pohon klasifikasi. Metode pemilihan pemilah pada pembentukan pohon klasifikasi perlu ditentukan yang terbaik diantara metode-metode yang mungkin, yaitu apakah dengan metode *test sample* yang membagi data sejumlah 43.544 unit RT menjadi data *learning* dan data *testing* dengan proporsi tertentu kemudian mengkombinasikannya, atau dengan metode *cross validation 10-fold* yang membagi data sejumlah 43.544 unit RT secara acak menjadi 10 bagian yang saling lepas dengan proporsi masing-masing kelas variabel respon relatif sama banyaknya. Selain itu juga perlu diperhatikan fungsi keheterogenan untuk menentukan pemilah terbaik, apakah digunakan ukuran fungsi indeks Gini atau indeks lain atau dalam hal ini dicobakan aturan indeks Twoing. Berikut adalah hasil dari pengolahan data rumah tangga miskin di kabupaten Jombang

dengan mencobakan metode-metode yang memungkinkan untuk dipilih tersebut agar dapat ditentukan metode pemilahan pembentukan pohon klasifikasi terbaik yaitu yang memberikan hasil ketepatan klasifikasi data *testing* tertinggi.

Tabel 4.3 Perbandingan Ketepatan Klasifikasi Hasil Beberapa Kemungkinan Metode Pemilahan Pembentuk Pohon Klasifikasi

	<i>Learning</i>	Jml. Data	<i>Testing</i>	Jml. Data	% ketepatan klasifikasi		Banyak simpul (Node)
					<i>L</i>	<i>T</i>	
Test Sample dengan Indeks Gini	95%	41.367	5%	2.177	76,9	36,2	7.115
	90%	39.190	10%	4.354	40,5	33,2	23
	85%	37.012	15%	6.532	34,5	28,3	6
	80%	34.835	20%	8.709	38,2	33,4	7
	75%	32.658	25%	10.886	37,4	36,1	3
	70%	30.481	30%	13.063	35,9	34,1	8
Test Sample dengan Indeks Twoing	95%	41.367	5%	2.177	67,4	36,5	3.781
	90%	39.190	10%	4.354	68,1	37,1	3.686
	85%	37.012	15%	6.532	37,7	29,6	6
	80%	34.835	20%	8.709	38,2	33,4	7
	75%	32.658	25%	10.886	40,8	33	80
	70%	30.481	30%	13.063	43,5	37,2	119
Cross Validation (CV) 10- fold	Indeks Gini				43,4	39,7	134
	Indeks Twoing				41	39,7	53

ket: L= *Learning*, T = *Testing*

Pada penelitian ini, *test sample* dicoba dengan enam kombinasi data *learning* dan *testing* yaitu 95 persen : 5 persen, 90 persen : 10 persen dan seterusnya sampai 70 persen : 30 persen dengan hasil yang diperoleh sebagaimana ditampilkan Tabel 4.2. Masing-masing metode dicoba dengan fungsi keheterogenan indeks Gini dan indeks Twoing. Dari Tabel 4.2 diperoleh informasi bahwa persentase ketepatan klasifikasi data *testing* dengan metode pemilahan CV menghasilkan nilai yang lebih besar

daripada nilai-nilai ketepatan klasifikasi dengan metode *test sample* pada semua kombinasi data. Ketepatan klasifikasi CV dengan indeks Gini maupun indeks Twoing bernilai sama yaitu 39,7 namun jumlah simpul terminal pohon klasifikasi yang terbentuk jika digunakan indeks Twoing jumlahnya lebih sedikit (ada 53) daripada jika digunakan indeks Gini (ada 134). Maka berdasarkan konsep parsimony, dipilih metode CV dengan indeks Twoing untuk pembentukan pohon klasifikasi terhadap rumah tangga miskin di kabupaten Jombang tahun 2010 dikarenakan dapat memberikan informasi ketepatan klasifikasi yang lebih besar dengan jumlah simpul terminal yang lebih sederhana. Ketepatan klasifikasi data *testing* dijadikan sebagai dasar pemilihan metode pembentukan model pohon klasifikasi sebab dapat memberikan gambaran kebaikan pohon klasifikasi yang nantinya terbentuk untuk mengklasifikasikan data baru.

Pembentukan pohon klasifikasi pada mulanya berupa pohon klasifikasi maksimal yaitu pohon klasifikasi yang memiliki jumlah simpul paling banyak. Setiap pemilah pada suatu simpul hanya bergantung pada nilai yang berasal dari satu variabel prediktor dimana setiap variabel prediktor mempunyai sejumlah kemungkinan pemilah, tergantung skala data variabel prediktor tersebut sebagaimana dijelaskan dalam subbab 2.1. Dari hasil pengolahan data, terbentuk pohon klasifikasi maksimal dengan simpul terminal sebanyak 7.842 simpul. Jika digambarkan, pohon klasifikasi maksimal tersebut sangat besar dan lebar atau kompleks karena memiliki simpul terminal sangat banyak. Diantara kedelapan belas variabel prediktor yang diteliti, luas kavling bangunan (X_2) mempunyai skor variabel penting (*important variable*) tertinggi yaitu 100 sebagaimana ditampilkan Tabel 4.4. Dengan demikian dapat dikatakan bahwa variabel prediktor luas kavling bangunan memiliki peranan utama atau merupakan faktor paling penting dalam pembentukan pohon klasifikasi maksimal terhadap klasifikasi rumah tangga miskin di kabupaten Jombang tahun 2010. Variabel prediktor yang terpenting berikutnya dalam pengklasifikasian ini adalah luas lantai (X_3) dengan skor 89, 94. Sementara itu, enam belas variabel prediktor lainnya memiliki skor di bawah 50.

Tabel 4.4 Variabel Penting Pembentukan Pohon Klasifikasi Maksimal

Variabel	Skor	
X ₂	100.00	
X ₃	84.94	
X ₁₇	41.68	
X ₁₈	37.28	
X ₆	27.40	
X ₇	26.54	
X ₁₀	23.44	
X ₁₆	22.86	
X ₁₁	19.29	
X ₁₂	18.72	
X ₁	18.14	
X ₈	17.95	
X ₁₃	13.53	
X ₅	13.01	
X ₉	11.75	
X ₁₅	10.71	
X ₁₄	7.16	
X ₄	6.52	

Tabel 4.5 Klasifikasi Data *Learning* oleh Pohon Klasifikasi Maksimal

Kelas Aktual	Kelas Prediksi		
	RTSM	RTM	RTHM
RTSM	6.706	295	223
RTM	2.311	13.317	1.917
RTHM	2.245	2.997	13.533

Keakuratan klasifikasi pohon maksimal yang dihasilkan dari data *learning* dapat dihitung berdasarkan Tabel 4.5. Hasil perhitungan $1 - \text{APER}$ sesuai persamaan (2.6) diperoleh sebesar 77,062 persen. Artinya bahwa pohon klasifikasi maksimal mampu mengklasifikasikan pengamatan ke dalam kelas-kelas variabel respon dengan tepat sebesar 77,062 persen. Dalam hal ini, terjadi kesalahan prediksi klasifikasi kelas pengamatan pada masing-masing kelas yaitu ada sebanyak 295 rumah tangga miskin yang secara aktual termasuk kelas RTSM salah diklasifikasikan sebagai kelas RTM dan 223 pengamatan sebagai kelas RTHM. Kemudian ada 2.311 rumah tangga miskin yang secara aktual termasuk kelas RTM salah diklasifikasikan sebagai RTSM dan 1.917 sebagai

kelas RTHM. Serta ada 2.245 rumah tangga miskin yang secara aktual termasuk kelas RTHM salah diklasifikasikan sebagai kelas RTSM dan 2.997 sebagai kelas RTM.

Tabel 4.6 menunjukkan hasil klasifikasi data *testing* oleh pohon klasifikasi maksimal. Keakuratan klasifikasi yang dihasilkan dengan menghitung 1- APER yaitu sebesar 39,068 persen. Artinya bahwa pohon klasifikasi maksimal memiliki keakuratan hasil prediksi pengamatan ke dalam kelas-kelas variabel respon sebesar 39,068 persen. Terdapat kesalahan prediksi klasifikasi kelas pengamatan pada masing-masing kelas yaitu ada sebanyak 2.604 rumah tangga miskin yang secara aktual termasuk kelas RTSM salah diprediksi sebagai kelas RTM dan 2.259 pengamatan sebagai kelas RTHM. Kemudian ada 4.587 rumah tangga miskin yang secara aktual termasuk kelas RTM salah diprediksi sebagai RTSM dan 6.177 sebagai kelas RTHM. Serta ada 4.292 rumah tangga miskin yang secara aktual termasuk kelas RTHM salah diprediksi sebagai kelas RTSM dan 6.613 sebagai kelas RTM.

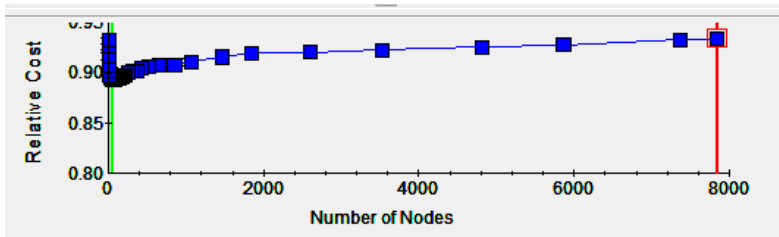
Tabel 4.6 Klasifikasi Data *Testing* oleh Pohon Klasifikasi Maksimal

Kelas Aktual	Kelas Prediksi		
	RTSM	RTM	RTHM
RTSM	2.361	2.604	2.259
RTM	4.587	6.781	6.177
RTHM	4.292	6.613	7.870

Pohon klasifikasi maksimal dengan 7.842 simpul terminal memiliki kedalaman pohon (*depth*) yang sangat tinggi. Dengan kata lain pohon klasifikasi maksimal yang terbentuk memiliki ukuran yang sangat besar. Nilai kompleksitas yang dihasilkan yaitu 0,000 dengan nilai penduga pengganti (*resubstitution relative cost*) sebesar 0,296 dan biaya kesalahan (*cross validated relative cost*) sebesar $0,934 \pm 0,004$ (antara 0,930 sampai 0,938). Kompleksitas senilai 0,000 menunjukkan struktur data yang kompleks dan pada umumnya cenderung bersifat *overfit*. Pohon klasifikasi yang kompleks nantinya juga akan menambah kesulitan peneliti dalam hal interpretasi hasil klasifikasi. Dengan demikian perlu dilakukan pemangkasan pohon (*pruning*) untuk

mendapatkan ukuran pohon klasifikasi yang layak dengan tingkat kompleksitas struktur data yang lebih kecil.

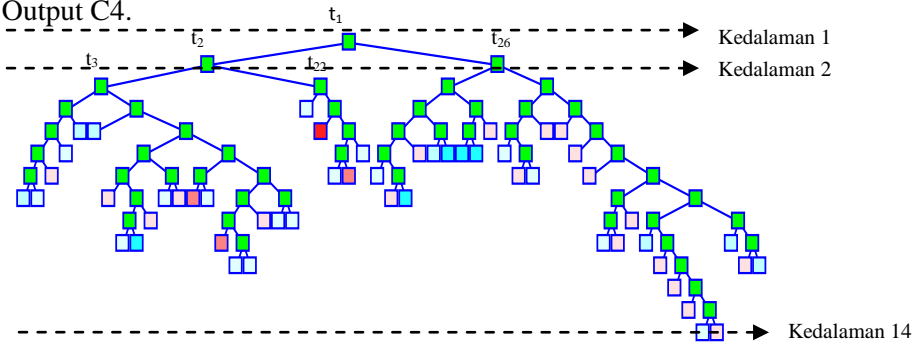
Pemangkasan pohon dilakukan dengan metode *cross validation 10-fold estimate*. Hasil pengolahan data menunjukkan terbentuknya pohon-pohon klasifikasi dengan jumlah simpul terminal tertentu dan juga nilai *relative cost* tertentu. Gambar 4.13 menampilkan adanya perbedaan nilai *relative cost* yang dihasilkan oleh pohon klasifikasi maksimal dengan pohon klasifikasi yang dianggap optimal. Nilai *relative cost* yang dihasilkan oleh pohon klasifikasi maksimal dengan jumlah simpul terminal 7.842 simpul adalah sebesar 0,934 (garis merah). Nilai tersebut lebih besar daripada nilai *relative cost* pohon klasifikasi optimal yang memiliki simpul terminal sebanyak 53 simpul yaitu sebesar 0,894 (garis hijau). Nilai kompleksitas pohon klasifikasi optimal yang diperoleh sebesar 0,000364 dan biaya kesalahan sebesar 0.894 ± 0.004 atau antara 0,890 sampai 0,898. Hasil selengkapnya bisa dilihat di Lampiran Output B2. Karena nilai *relative cost* pohon klasifikasi optimal lebih kecil maka pohon klasifikasi optimal dipilih sebagai pohon yang layak untuk pohon klasifikasi rumah tangga miskin di kabupaten Jombang tahun 2010.



Gambar 4.13 Plot *relative cost* dengan jumlah simpul tertentu.

Pemangkasan pohon klasifikasi maksimal menghasilkan pohon klasifikasi optimal dengan jumlah simpul terminal sebanyak 53 simpul dan kedalaman pohon 14. Simpul terminal pohon sebanyak 53 pada Gambar 4.14 ditunjukkan dengan warna degradasi antara merah, pink, putih, dan kebiruan. Perbedaan warna tersebut ada kaitannya dengan pemberian label kelas masing-masing simpul terminal. Simpul terminal yang berwarna

merah diberi label kelas 1 (RTSM). Jika warna simpul terminal semakin merah kuat atau pekat, hal itu menunjukkan rumah tangga miskin anggota simpul terminal tersebut yang termasuk kelas RTSM persentasenya mendekati 100 persen. Jika warna merah semakin memudar dan semakin menuju warna putih, hal itu menunjukkan bahwa rumah tangga miskin anggota simpul tersebut yang termasuk kelas RTSM persentasenya menurun dan bahkan lebih kecil daripada persentase rumah tangga miskin yang juga anggota simpul tersebut tapi termasuk kelas RTM. Simpul terminal tersebut diberi label kelas 2 (RTM). Sementara jika warna simpul terminal semakin menuju warna biru, hal itu menunjukkan bahwa rumah tangga miskin anggota simpul tersebut yang termasuk kelas RTHM persentasenya mencapai 70 persen ke atas. Simpul terminal tersebut diberi label kelas 3 (RTHM). Pohon klasifikasi optimal yang lebih jelas dan rinci dengan penjelasan masing-masing pemilah simpul dan banyaknya anggota masing-masing simpul dapat dilihat pada Lampiran Output C4.



Gambar 4.14 Topologi pohon klasifikasi optimal.

Variabel terpenting dalam pembentukan pohon klasifikasi optimal yang mengklasifikasikan rumah tangga miskin di kabupaten Jombang tahun 2010 berturut-berturut yaitu X_{17} (penghasilan rata-rata per bulan) dengan skor 100, X_{18} (aset yang dimiliki dalam satuan Rupiah) dengan skor 80,29, X_3 (luas lantai) dengan skor 61,73, dan X_8 (tempat pembuangan akhir tinja) dengan skor 60,18. Variabel prediktor yang lain mendapatkan skor di bawah itu yang menunjukkan semakin menurunnya

pengaruh variabel tersebut dalam menentukan klasifikasi rumah tangga miskin melalui pemilahan simpul-simpul pembentuk pohon klasifikasi optimal. Secara lengkap skor per variabel ditampilkan dalam Tabel 4.7. Sementara itu, variabel X_{17} (penghasilan) terpilih sebagai pemilah simpul utama karena mampu memberikan nilai penurunan keheterogenan tertinggi pada simpul 1 sebagaimana ditunjukkan dalam Tabel 4.8.

Tabel 4.7 Skor Variabel Prediktor Penting Pembentuk Pohon Klasifikasi Optimal

Variabel	Skor	
X_{17}	100.00	
X_{18}	80.29	
X_3	61.73	
X_8	60.18	
X_{15}	41.66	
X_5	37.02	
X_{10}	26.34	
X_2	19.14	
X_7	18.64	
X_{12}	7.71	
X_{11}	7.65	
X_{16}	6.58	
X_{14}	6.39	
X_9	5.08	
X_6	3.11	
X_{13}	0.00	
X_1	0.00	
X_4	0.00	

Tabel 4.8 Kemungkinan Pemilah Simpul 1 Pohon Klasifikasi Optimal

Variabel Prediktor	Pemilah	Penurunan Kehlerogenan
X_{18}	0	0,008
X_{15}	2; 3; 4	0,005
X_{16}	1; 4; 6	0,004
X_5	1 ; 3	0,002
X_{13}	0,5	0,002
$X_{17}^{*)}$	282.000	0,010

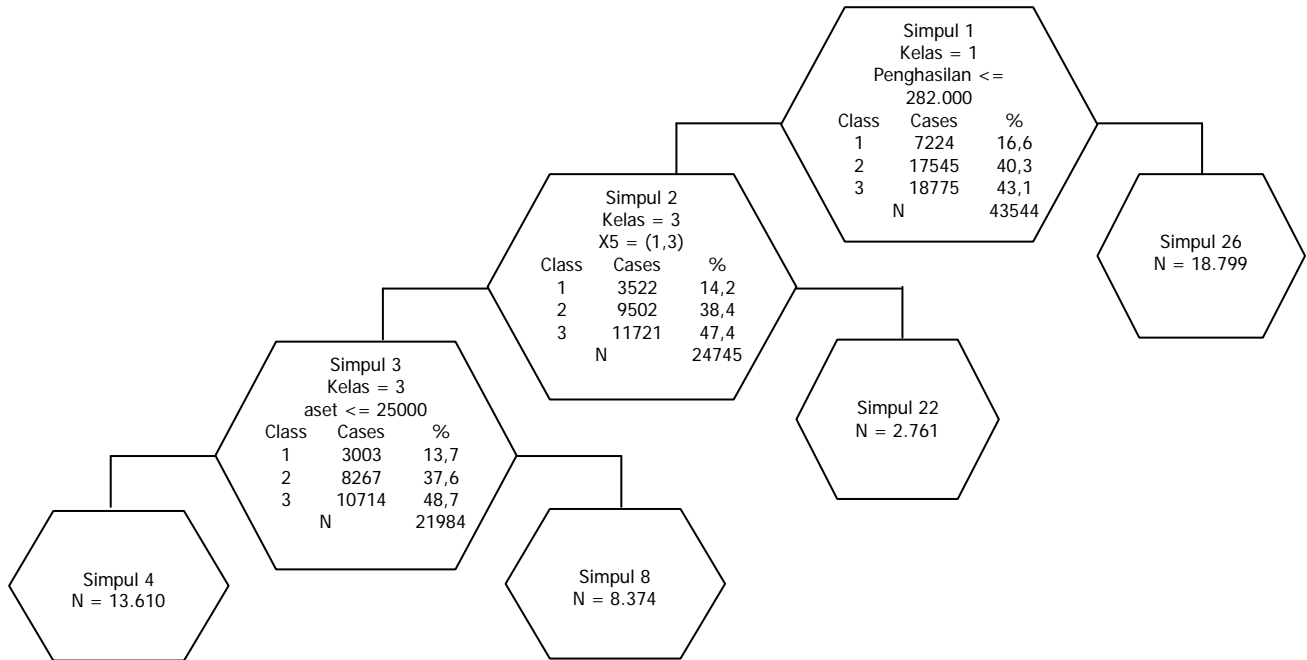
^{*)} Pemilah terpilih

Variabel X_{17} (penghasilan) memilah simpul utama (simpul 1) menjadi simpul kiri dan simpul kanan dengan ketentuan rumah

tangga miskin yang berpenghasilan \leq Rp 282.000,- akan dipilah menjadi simpul kiri (simpul 2) sedangkan rumah tangga miskin yang berpenghasilan $>$ Rp 282.000,- akan dipilah menjadi simpul kanan (simpul 26). Diperoleh hasil bahwa ada sebanyak 24.745 rumah tangga miskin memiliki penghasilan \leq Rp 282.000,- yang menjadi anggota simpul kiri (simpul 2) dan sisanya yaitu 18.799 rumah tangga miskin berpenghasilan $>$ Rp 282.000,- menjadi anggota simpul kanan (simpul 26). Lebih jelasnya bisa dilihat di Lampiran Output C2 mengenai informasi pemilahan simpul.

Simpul 2 yang beranggotakan 24.745 rumah tangga miskin dengan penghasilan \leq Rp 282.000,- selanjutnya dipilah menjadi simpul baru kiri dan kanan menurut jenis dinding terluas (X_5). Jika jenis dinding terluas rumah tangga miskin yaitu tembok ataupun bambu (kategori 1 ataupun 3) maka rumah tangga miskin tersebut akan dipilah menjadi simpul kiri baru (simpul 3). Namun jika jenis dinding terluasnya adalah kayu (kategori 2) maka akan dipilah menjadi anggota simpul kanan baru (simpul 22). Diantara 24.745 rumah tangga miskin anggota simpul 2, diperoleh hasil ada sebanyak 21.984 rumah tangga miskin yang menjadi anggota simpul 3 dengan karakteristik penghasilan \leq Rp 282.000,- dan jenis dinding terluas yaitu tembok atau bambu (kategori 1 atau 3). Sisanya ada 2.761 rumah tangga miskin yang menjadi anggota simpul 22 dengan karakteristik penghasilan $>$ Rp 282.000,- dan jenis dinding terluas yaitu kayu (kategori 2).

Simpul 3 yang beranggotakan 21.984 rumah tangga miskin dengan jenis dinding terluas yaitu tembok atau bambu (kategori 1 atau 3) selanjutnya dipilah menjadi simpul baru kiri dan kanan menurut besarnya aset yang dimiliki. Jika aset yang dimiliki rumah tangga miskin \leq Rp 25.000,- maka rumah tangga miskin tersebut akan dipilah ke simpul kiri baru (simpul 4). Jika tidak maka akan dipilah ke simpul kanan baru (simpul 8). Diperoleh hasil bahwa diantara 21.984 rumah tangga miskin anggota simpul 3, ada sebanyak 13.610 rumah tangga miskin yang memiliki aset \leq Rp 25.000,- sehingga dipilah ke simpul kiri baru (simpul 4) dan sisanya ada 8.374 rumah tangga miskin dipilah ke simpul 8.



Gambar 4.15 Potongan struktur pohon klasifikasi optimal untuk visualisasi interpretasi.

Pada Gambar 4.15 diberikan visualisasi potongan struktur pohon klasifikasi optimal untuk penjelasan pemilahan simpul yang dijelaskan di atas, mulai dari pemilahan simpul utama (simpul 1) sampai pemilahan simpul 3 menjadi simpul 4 dan simpul 8. Dengan demikian diharapkan agar interpretasi struktur pohon klasifikasi yang terbentuk bisa lebih mudah dipahami secara nyata atau jelas. Struktur pohon klasifikasi optimal lebih rinci dengan penjelasan masing-masing pemilah simpul, banyaknya anggota, dan informasi label kelas masing-masing simpul dapat dilihat pada Lampiran Output C4.

Suatu simpul akan terus dipilah menjadi simpul anak baru (kiri dan kanan) sesuai prosedur *binary recursive partitioning*, sampai simpul tersebut telah dianggap memiliki anggota yang homogen atau jika simpul tersebut hanya memiliki 1 anggota pengamatan (dalam hal ini 1 rumah tangga miskin) maka simpul akan menjadi simpul terminal dan tidak akan dipilah lagi. Pohon klasifikasi optimal yang terbentuk terdiri atas 53 simpul terminal sebagaimana telah ditunjukkan pada Gambar 4.14. Masing-masing simpul terminal tersebut memiliki karakteristik tertentu dan diprediksi sebagai kelas variabel respon tertentu sesuai dengan label kelas yang diberikan. Berdasarkan hasil penelusuran 53 simpul terminal pohon klasifikasi optimal tersebut, berikut ini diberikan rangkuman pengklasifikasian kelas rumah tangga miskin menurut indikasi kesamaan label kelas setiap simpul terminal.

Tabel 4.9 Kelas Rumah Tangga Miskin pada Masing-masing Simpul Terminal

Kelas	simpul	Persentase	simpul	Persentase	simpul	Persentase
	terminal ke-		terminal ke-		terminal ke-	
RTSM (23 simpul terminal)	3	22,1	24	48,4	41	22,5
	7	32,9	27	27,6	44	25,9
	10	29,8	29	19,2	46	31,3
	12	22,7	34	32,1	47	23,5
	13	46,6	36	23,2	48	19,9
	15	42,9	38	21,3	50	26,2
	18	24,4	39	30,6	52	26,4
	22	71,2	40	25,5		

Tabel 4.9 Lanjutan

Kelas	simpul terminal ke-	Persentase	simpul terminal ke-	Persentase	simpul terminal ke-	Persentase
RTM (18 simpul terminal)	2	42	23	64,4	42	46,5
	8	44,1	25	46,4	43	44,6
	11	44	28	57,3	45	46,4
	16	49,8	30	46,5	49	47,9
	19	45,6	32	58	51	46
	21	46	37	41,5	53	48,7
RTHM (12 simpul terminal)	1	52,7	9	58,5	26	45,9
	4	51,3	14	48,9	31	73,9
	5	52,9	17	47,1	33	55,3
	6	58,9	20	48,6	35	50,8

Penelusuran struktur pohon klasifikasi optimal terhadap simpul terminal dapat memberikan informasi tentang karakteristik kelas simpul terminal tersebut. Berikut adalah karakteristik dari simpul terminal dengan persentase tertinggi untuk masing-masing kelas.

Tabel 4.10 Karakteristik Kelas Rumah Tangga Miskin Menurut Persentase Kelas Tertinggi Simpul Terminal

Kelas	Karakteristik
(1) RTSM	luas lantai (X_3) $\leq 9,5 \text{ m}^2$, tempat pembuangan air tinja (X_8) di lubang tanah, jenis dinding terluas (X_5) dari kayu, dan penghasilan (X_{17}) $\leq \text{Rp } 282.000,-$
(2) RTM	asset yang dimiliki (X_{18}) $\leq \text{Rp } 65.000$, sumber air minum air (X_{10}) dalam kemasan atau mata air atau air sungai, luas lantai (X_3) $> 9,5 \text{ m}^2$, tempat pembuangan air tinja (X_8) di lubang tanah, jenis dinding terluas (X_5) dari kayu, dan penghasilan (X_{17}) $\leq \text{Rp } 282.000,-$
(3) RTHM	kepemilikan asset (X_{18}) $\leq \text{Rp } 1.137.500$, intensitas makan daging susu ayam per minggu (X_{12}) ≤ 2 kali, pengobatan (X_{15}) dilakukan dengan bantuan praktik dokter atau paramedis atau pengobatan tradisional, dan penghasilan (X_{17}) $> \text{Rp } 282.000,-$

Keakuratan klasifikasi pohon optimal yang dihasilkan dari data *learning* dapat dihitung berdasarkan Tabel 4.11. Hasil perhitungan 1 – APER diperoleh sebesar 40,986 persen. Artinya

bahwa pohon klasifikasi optimal mampu mengklasifikasikan suatu rumah tangga miskin ke dalam kelas-kelas status kemiskinan rumah tangga miskin (RTSM, RTM atau RTHM) dengan tepat sebesar 40,986 persen. Terjadi kesalahan prediksi klasifikasi kelas pengamatan pada masing-masing kelas yaitu ada sebanyak 1.205 rumah tangga miskin yang secara aktual termasuk kelas RTSM salah diklasifikasikan sebagai kelas RTM dan 2.422 pengamatan sebagai kelas RTHM. Kemudian ada 6.250 rumah tangga miskin yang secara aktual termasuk kelas RTM salah diklasifikasikan sebagai RTSM dan 7.138 sebagai kelas RTHM. Serta ada 4.995 rumah tangga miskin yang secara aktual termasuk kelas RTHM salah diklasifikasikan sebagai kelas RTSM dan 3.687 sebagai kelas RTM.

Tabel 4.11 Klasifikasi Data *Learning* oleh Pohon Klasifikasi Optimal

Kelas Aktual	Kelas Prediksi		
	RTSM	RTM	RTHM
RTSM	3.597	1.205	2.422
RTM	6.250	4.157	7.138
RTHM	4.995	3.687	10.093

Pohon klasifikasi optimal yang terbentuk perlu divalidasi untuk mengetahui apakah pohon klasifikasi tersebut layak dan dapat digunakan untuk mengklasifikasi data baru. Keakuratan klasifikasi data *testing* yang dihasilkan pada Tabel 4.12 dengan menghitung 1- APER adalah sebesar 39,654 persen. Artinya bahwa pohon klasifikasi optimal memiliki keakuratan hasil prediksi suatu rumah tangga miskin termasuk ke dalam salah satu kelas variabel respon sebesar 39,654 persen. Terdapat kesalahan prediksi klasifikasi masing-masing kelas yaitu ada sebanyak 1.346 rumah tangga miskin yang secara aktual termasuk kelas RTSM salah diprediksi sebagai kelas RTM dan 2.606 pengamatan sebagai kelas RTHM. Kemudian ada 6.102 rumah tangga miskin yang secara aktual termasuk kelas RTM salah diprediksi sebagai RTSM dan 7.595 sebagai kelas RTHM. Serta ada 5.111 rumah tangga miskin yang secara aktual termasuk kelas RTHM salah diprediksi sebagai kelas RTSM dan 3.517 sebagai kelas RTM.

Tabel 4.12 Klasifikasi Data *Testing* oleh Pohon Klasifikasi Optimal

Kelas Aktual	Kelas Prediksi		
	RTSM	RTM	RTHM
RTSM	3.272	1.346	2.606
RTM	6.102	3.848	7.595
RTHM	5.111	3.517	10.147

Hasil keakuratan klasifikasi pohon maksimal dengan pohon optimal jika dibandingkan pada Tabel 4.13, maka dapat diketahui bahwa untuk data *learning*, keakuratan klasifikasi pohon maksimal lebih tinggi daripada pohon optimal. Hal ini dikarenakan pohon klasifikasi maksimal memiliki simpul yang paling banyak dengan melibatkan lebih banyak variabel prediktor sebagai pemilah simpul sehingga kemungkinan untuk mengklasifikasikan data dengan tepat akan cenderung lebih besar. Sementara untuk data *testing*, keakuratan klasifikasi pohon optimal sedikit lebih tinggi dibanding pohon maksimal, namun kedua nilai tersebut sama-sama dikatakan masih relatif rendah sebab tidak mencapai 50 persen.

Tabel 4.13 Perbandingan Keakuratan Klasifikasi Pohon Maksimal dan Pohon Optimal

Pohon Klasifikasi	% Ketepatan Klasifikasi	
	<i>Learning</i>	<i>Testing</i>
Pohon Maksimal	77,062	39,068
Pohon Optimal	40,986	39,654

Penelitian sebelumnya yang juga menjadi acuan penelitian ini yaitu klasifikasi kemiskinan di Kabupaten Jombang dengan metode *ensemble* CART oleh Muttaqin (2013), menghasilkan akurasi prediksi pohon klasifikasi CART yang sangat rendah juga untuk kelas Rumah tangga Sangat Miskin, yaitu hanya sebesar 5,02 persen. Jika dikaitkan antara variabel terpenting pohon klasifikasi CART, yang menentukan pengklasifikasian kelas rumah tangga miskin, dengan dasar pembagian kategori atau kelas rumah tangga miskin memang ditemukan adanya ketidaksesuaian. Baik pada penelitian ini maupun penelitian Muttaqin, pembagian kelas rumah tangga miskin menjadi RTSM, RTM, dan RTHM mengacu pada ketetapan BPS yang didasarkan pada

besarnya pengeluaran rumah tangga miskin per bulan sedangkan dari hasil pohon klasifikasi CART variabel terpenting yang mempengaruhi pengklasifikasian rumah tangga miskin adalah variabel penghasilan per bulan. Demikian diduga menjadi salah satu sebab mengapa keakuratan klasifikasi yang dihasilkan relatif rendah. Namun tidak menutup kemungkinan jika hal tersebut disebabkan karena adanya pengaruh faktor lain yang tidak diperhatikan dalam penelitian ini. Jadi, dapat dikatakan bahwa untuk pengklasifikasian rumah tangga miskin di kabupaten Jombang tahun 2010 ke dalam kelas RTSM, RTM atau RTHM dalam hal ini ditentukan oleh besarnya penghasilan rata-rata per bulan sebagai variabel terpenting dalam klasifikasi dengan keakuratan klasifikasi sebesar 39,654 persen.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan uraian dan pembahasan yang telah dilakukan, diperoleh kesimpulan bahwa diantara 73.720 rumah tangga miskin di Kabupaten Jombang tahun 2010, sebanyak 11.676 rumah tangga atau sekitar 15,8 persen termasuk kelas RTSM, sebanyak 29.156 rumah tangga (39,6 persen) termasuk RTM, dan paling banyak yaitu 32.888 rumah tangga atau sekitar 44,6 persen termasuk kelas RTHM. Pohon klasifikasi yang layak untuk mengklasifikasikan kelas status kemiskinan rumah tangga miskin di Kabupaten Jombang yaitu.

- a. Pohon klasifikasi menggunakan metode pemilihan pemilah indeks Twoing dengan penduga *Cross Validation 10-fold* yang memiliki 53 simpul terminal dan kedalaman pohon 14. Data sampel *learning* secara keseluruhan tepat diklasifikasikan pohon klasifikasi sebesar 40,986 persen dan akurasi prediksi data *testing* sebesar 39,654 persen.
- b. Variabel terpenting dalam menentukan status kemiskinan suatu rumah tangga miskin pada penelitian ini yaitu besarnya penghasilan rata-rata per bulan dalam satuan Rupiah dengan skor tingkat kepentingan variabel sebesar 100.

Karakteristik RTSM menurut penelusuran hasil simpul terminal pohon klasifikasi dengan prosentase tertinggi yaitu luas lantai bangunan tempat tinggal $\leq 9,5 \text{ m}^2$, tempat pembuangan air tinja di lubang tanah, jenis dinding terluas dari kayu, dan penghasilan $\leq \text{Rp } 282.000,-$. Sementara karakteristik untuk RTM yaitu asset yang dimiliki $\leq \text{Rp } 65.000$, sumber air minum air dalam kemasan atau mata air atau air sungai, luas lantai bangunan tempat tinggal $> 9,5 \text{ m}^2$, tempat pembuangan air tinja di lubang tanah, jenis dinding terluas dari kayu, dan penghasilan $\leq \text{Rp } 282.000,-$. Sedangkan karakteristik untuk RTHM yaitu kepemilikan asset $\leq \text{Rp } 1.137.500$, intensitas makan daging susu ayam per minggu ≤ 2 kali, pengobatan dilakukan dengan bantuan praktik dokter atau paramedis atau pengobatan tradisional, dan penghasilan $> \text{Rp } 282.000,-$.

5.2 Saran

Sebagai saran untuk penelitian berikutnya adalah data yang akan dianalisis perlu disiapkan dengan benar apalagi jika data tersebut jumlahnya sangat besar dan memuat banyak variabel dengan skala campuran agar hasil-hasil yang mungkin tidak logis bisa dihindari dan ketepatan klasifikasi yang dihasilkan lebih tinggi atau lebih baik. Untuk mendapatkan nilai ketepatan klasifikasi yang lebih tinggi mungkin bisa dicobakan alternatif metode klasifikasi *ensemble* CART yang bisa digunakan untuk data rumah tangga miskin di kabupaten Jombang tahun 2010 ini.

Lampiran A. Data Penelitian

No	Y	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	X ₁₁	X ₁₂	X ₁₃	X ₁₄	X ₁₅	X ₁₆	X ₁₇	X ₁₈
1	2	1	24	24	2	1	6	4	3	2	4	5	1	1	3	1	1	0	0
2	3	1	90	12	2	1	3	2	3	1	4	2	1	1	3	1	1	0	0
3	3	1	18	15	2	1	2	1	1	1	3	3	1	2	3	1	2	700000	1000000

73.720	2	1	28	0	2	2	6	4	4	4	4	5	1	1	3	1	1	0	0

Y	Status kemiskinan Rumah Tangga	X ₇	Fasilitas tempat buang air besar (jamban)	X ₁₄	Intensitas makan per hari
X ₁	Status kepemilikan bangunan	X ₈	Tempat pembuangan akhir tinja	X ₁₅	Tempat Pengobatan
X ₂	Luas kavling (m ²)	X ₉	Sumber penerangan utama	X ₁₆	Ijazah terakhir kepala keluarga
X ₃	Luas lantai (m ²)	X ₁₀	Sumber air minum	X ₁₇	Penghasilan rata-rata tiap bulan (Rp)
X ₄	Jenis atap rumah terluas	X ₁₁	Bahan bakar memasak	X ₁₈	Kepemilikan asset dalam rumah (Rp)
X ₅	Jenis dinding terluas	X ₁₂	Intensitas konsumsi daging susu ayam per minggu		
X ₆	Jenis lantai terluas	X ₁₃	Intensitas membeli pakaian satu set per tahun		

Output A1. Ketepatan Klasifikasi 95% : 5% dengan Indeks Gini

=====				
TEST SAMPLE CLASSIFICATION TABLE				
=====				
Actual Class	Predicted Class			Actual Total
	1	2	3	
-----	-----	-----	-----	-----
1	93.00	212.00	176.00	481.00
2	217.00	396.00	350.00	963.00
3	162.00	272.00	299.00	733.00
-----	-----	-----	-----	-----
PRED. TOT.	472.00	880.00	825.00	2177.00
CORRECT	0.193	0.411	0.408	
SUCCESS IND.	-0.028	-0.031	0.071	
TOT. CORRECT	0.362			

=====				
TEST SAMPLE CLASSIFICATION PROBABILITY TABLE				
=====				
Actual Class	Predicted Class			Actual Total
	1	2	3	
-----	-----	-----	-----	-----
1	0.193	0.441	0.366	1.000
2	0.225	0.411	0.363	1.000
3	0.221	0.371	0.408	1.000

=====				
LEARNING SAMPLE CLASSIFICATION TABLE				
=====				
Actual Class	Predicted Class			Actual Total
	1	2	3	
-----	-----	-----	-----	-----
1	6266.00	259.00	218.00	6743.00
2	2052.00	12270.00	2260.00	16582.00
3	2261.00	2510.00	13271.00	18042.00
-----	-----	-----	-----	-----
PRED. TOT.	10579.00	15039.00	15749.00	41367.00
CORRECT	0.929	0.740	0.736	
SUCCESS IND.	0.766	0.339	0.299	
TOT. CORRECT	0.769			

=====				
LEARNING SAMPLE CLASSIFICATION PROBABILITY TABLE				
=====				
Actual Class	Predicted Class			Actual Total
	1	2	3	
-----	-----	-----	-----	-----
1	0.929	0.038	0.032	1.000
2	0.124	0.740	0.136	1.000
3	0.125	0.139	0.736	1.000

Output A2. Ketepatan Klasifikasi 95% : 5% dengan Indeks Twoing

TEST SAMPLE CLASSIFICATION TABLE

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	126.00	230.00	125.00	481.00
2	244.00	417.00	302.00	963.00
3	189.00	293.00	251.00	733.00
PRED. TOT.	559.00	940.00	678.00	2177.00
CORRECT	0.262	0.433	0.342	
SUCCESS IND.	0.041	-0.009	0.006	
TOT. CORRECT	0.365			

TEST SAMPLE CLASSIFICATION PROBABILITY TABLE

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	0.262	0.478	0.260	1.000
2	0.253	0.433	0.314	1.000
3	0.258	0.400	0.342	1.000

LEARNING SAMPLE CLASSIFICATION TABLE

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	5861.00	439.00	443.00	6743.00
2	2805.00	10541.00	3236.00	16582.00
3	2819.00	3747.00	11476.00	18042.00
PRED. TOT.	11485.00	14727.00	15155.00	41367.00
CORRECT	0.869	0.636	0.636	
SUCCESS IND.	0.706	0.235	0.200	
TOT. CORRECT	0.674			

LEARNING SAMPLE CLASSIFICATION PROBABILITY TABLE

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	0.869	0.065	0.066	1.000
2	0.169	0.636	0.195	1.000
3	0.156	0.208	0.636	1.000

Output A3. Ketepatan Klasifikasi 90% : 10% dengan Indeks Gini

=====

TEST SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	222.00	96.00	426.00	744.00
2	529.00	333.00	1190.00	2052.00
3	419.00	250.00	889.00	1558.00
PRED. TOT.	1170.00	679.00	2505.00	4354.00
CORRECT	0.298	0.162	0.571	
SUCCESS IND.	0.128	-0.309	0.213	
TOT. CORRECT	0.332			

=====

TEST SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	0.298	0.129	0.573	1.000
2	0.258	0.162	0.580	1.000
3	0.269	0.160	0.571	1.000

=====

LEARNING SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	3121.00	674.00	2685.00	6480.00
2	5726.00	1979.00	7788.00	15493.00
3	4553.00	1886.00	10778.00	17217.00
PRED. TOT.	13400.00	4539.00	21251.00	39190.00
CORRECT	0.482	0.128	0.626	
SUCCESS IND.	0.316	-0.268	0.187	
TOT. CORRECT	0.405			

=====

LEARNING SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	0.482	0.104	0.414	1.000
2	0.370	0.128	0.503	1.000
3	0.264	0.110	0.626	1.000

Output A4. Ketepatan Klasifikasi 90% : 10% dengan Indeks Twoing

```
=====
TEST SAMPLE CLASSIFICATION TABLE
=====
```

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	215.00	278.00	251.00	744.00
2	503.00	804.00	745.00	2052.00
3	394.00	567.00	597.00	1558.00
PRED. TOT.	1112.00	1649.00	1593.00	4354.00
CORRECT	0.289	0.392	0.383	
SUCCESS IND.	0.118	-0.079	0.025	
TOT. CORRECT	0.371			

```
-----
```

```
=====
TEST SAMPLE CLASSIFICATION PROBABILITY TABLE
=====
```

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	0.289	0.374	0.337	1.000
2	0.245	0.392	0.363	1.000
3	0.253	0.364	0.383	1.000

```
-----
```

```
=====
LEARNING SAMPLE CLASSIFICATION TABLE
=====
```

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	5647.00	437.00	396.00	6480.00
2	2577.00	10049.00	2867.00	15493.00
3	2703.00	3507.00	11007.00	17217.00
PRED. TOT.	10927.00	13993.00	14270.00	39190.00
CORRECT	0.871	0.649	0.639	
SUCCESS IND.	0.706	0.253	0.200	
TOT. CORRECT	0.681			

```
-----
```

```
=====
LEARNING SAMPLE CLASSIFICATION PROBABILITY TABLE
=====
```

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	0.871	0.067	0.061	1.000
2	0.166	0.649	0.185	1.000
3	0.157	0.204	0.639	1.000

```
-----
```

Output A5. Ketepatan Klasifikasi 85% : 15% dengan Indeks Gini

=====

TEST SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	609.00	0.00	531.00	1140.00
2	1452.00	0.00	1702.00	3154.00
3	999.00	0.00	1239.00	2238.00

PRED. TOT.	3060.00	0.00	3472.00	6532.00
CORRECT	0.534	0.000	0.554	
SUCCESS IND.	0.360	-0.483	0.211	
TOT. CORRECT	0.283			

=====

TEST SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	0.534	0.000	0.466	1.000
2	0.460	0.000	0.540	1.000
3	0.446	0.000	0.554	1.000

=====

LEARNING SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	4035.00	0.00	2049.00	6084.00
2	8369.00	0.00	6022.00	14391.00
3	7815.00	0.00	8722.00	16537.00

PRED. TOT.	20219.00	0.00	16793.00	37012.00
CORRECT	0.663	0.000	0.527	
SUCCESS IND.	0.499	-0.389	0.081	
TOT. CORRECT	0.345			

=====

LEARNING SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	0.663	0.000	0.337	1.000
2	0.582	0.000	0.418	1.000
3	0.473	0.000	0.527	1.000

Output A6. Ketepatan Klasifikasi 85% : 15% dengan Indeks Twoing

TEST SAMPLE CLASSIFICATION TABLE

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	563.00	0.00	577.00	1140.00
2	1358.00	0.00	1796.00	3154.00
3	865.00	0.00	1373.00	2238.00
PRED. TOT.	2786.00	0.00	3746.00	6532.00
CORRECT	0.494	0.000	0.613	
SUCCESS IND.	0.319	-0.483	0.271	
TOT. CORRECT	0.296			

TEST SAMPLE CLASSIFICATION PROBABILITY TABLE

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	0.494	0.000	0.506	1.000
2	0.431	0.000	0.569	1.000
3	0.387	0.000	0.613	1.000

LEARNING SAMPLE CLASSIFICATION TABLE

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	3354.00	0.00	2730.00	6084.00
2	6813.00	0.00	7578.00	14391.00
3	5947.00	0.00	10590.00	16537.00
PRED. TOT.	16114.00	0.00	20898.00	37012.00
CORRECT	0.551	0.000	0.640	
SUCCESS IND.	0.387	-0.389	0.194	
TOT. CORRECT	0.377			

LEARNING SAMPLE CLASSIFICATION PROBABILITY TABLE

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	0.551	0.000	0.449	1.000
2	0.473	0.000	0.527	1.000
3	0.360	0.000	0.640	1.000

Output A7. Ketepatan Klasifikasi 80% : 20% dengan Indeks Gini

=====

TEST SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	629.00	0.00	1012.00	1641.00
2	1276.00	0.00	2657.00	3933.00
3	852.00	0.00	2283.00	3135.00

PRED. TOT.	2757.00	0.00	5952.00	8709.00
CORRECT	0.383	0.000	0.728	
SUCCESS IND.	0.195	-0.452	0.368	
TOT. CORRECT	0.334			

=====

TEST SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	0.383	0.000	0.617	1.000
2	0.324	0.000	0.676	1.000
3	0.272	0.000	0.728	1.000

=====

LEARNING SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	3019.00	0.00	2564.00	5583.00
2	6182.00	0.00	7430.00	13612.00
3	5349.00	0.00	10291.00	15640.00

PRED. TOT.	14550.00	0.00	20285.00	34835.00
CORRECT	0.541	0.000	0.658	
SUCCESS IND.	0.380	-0.391	0.209	
TOT. CORRECT	0.382			

=====

LEARNING SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	0.541	0.000	0.459	1.000
2	0.454	0.000	0.546	1.000
3	0.342	0.000	0.658	1.000

Output A8. Ketepatan Klasifikasi 80% : 20% dengan Indeks Twoing

=====

TEST SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	629.00	0.00	1012.00	1641.00
2	1279.00	0.00	2654.00	3933.00
3	854.00	0.00	2281.00	3135.00
PRED. TOT.	2762.00	0.00	5947.00	8709.00
CORRECT	0.383	0.000	0.728	
SUCCESS IND.	0.195	-0.452	0.368	
TOT. CORRECT	0.334			

=====

TEST SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	0.383	0.000	0.617	1.000
2	0.325	0.000	0.675	1.000
3	0.272	0.000	0.728	1.000

=====

LEARNING SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	3021.00	0.00	2562.00	5583.00
2	6190.00	0.00	7422.00	13612.00
3	5353.00	0.00	10287.00	15640.00
PRED. TOT.	14564.00	0.00	20271.00	34835.00
CORRECT	0.541	0.000	0.658	
SUCCESS IND.	0.381	-0.391	0.209	
TOT. CORRECT	0.382			

=====

LEARNING SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	0.541	0.000	0.459	1.000
2	0.455	0.000	0.545	1.000
3	0.342	0.000	0.658	1.000

Output A9. Ketepatan Klasifikasi 75% : 25% dengan Indeks Gini

=====

TEST SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	706.00	0.00	1126.00	1832.00
2	1487.00	0.00	3058.00	4545.00
3	1286.00	0.00	3223.00	4509.00
PRED. TOT.	3479.00	0.00	7407.00	10886.00
CORRECT	0.385	0.000	0.715	
SUCCESS IND.	0.217	-0.418	0.301	
TOT. CORRECT	0.361			

=====

TEST SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	0.385	0.000	0.615	1.000
2	0.327	0.000	0.673	1.000
3	0.285	0.000	0.715	1.000

=====

LEARNING SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	2785.00	0.00	2607.00	5392.00
2	5682.00	0.00	7318.00	13000.00
3	4827.00	0.00	9439.00	14266.00
PRED. TOT.	13294.00	0.00	19364.00	32658.00
CORRECT	0.517	0.000	0.662	
SUCCESS IND.	0.351	-0.398	0.225	
TOT. CORRECT	0.374			

=====

LEARNING SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	0.517	0.000	0.483	1.000
2	0.437	0.000	0.563	1.000
3	0.338	0.000	0.662	1.000

Output A10. Ketepatan Klasifikasi 75% : 25% dengan Indeks Twoing

=====

TEST SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	949.00	253.00	630.00	1832.00
2	2116.00	750.00	1679.00	4545.00
3	1940.00	681.00	1888.00	4509.00
PRED. TOT.	5005.00	1684.00	4197.00	10886.00
CORRECT	0.518	0.165	0.419	
SUCCESS IND.	0.350	-0.252	0.005	
TOT. CORRECT	0.330			

=====

TEST SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	0.518	0.138	0.344	1.000
2	0.466	0.165	0.369	1.000
3	0.430	0.151	0.419	1.000

=====

LEARNING SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	3178.00	693.00	1521.00	5392.00
2	5481.00	2605.00	4914.00	13000.00
3	4601.00	2120.00	7545.00	14266.00
PRED. TOT.	13260.00	5418.00	13980.00	32658.00
CORRECT	0.589	0.200	0.529	
SUCCESS IND.	0.424	-0.198	0.092	
TOT. CORRECT	0.408			

=====

LEARNING SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	0.589	0.129	0.282	1.000
2	0.422	0.200	0.378	1.000
3	0.323	0.149	0.529	1.000

Output A11. Ketepatan Klasifikasi 70% : 30% dengan Indeks Gini

TEST SAMPLE CLASSIFICATION TABLE

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	1078.00	0.00	985.00	2063.00
2	2522.00	0.00	2646.00	5168.00
3	2454.00	0.00	3378.00	5832.00
PRED. TOT.	6054.00	0.00	7009.00	13063.00
CORRECT	0.523	0.000	0.579	
SUCCESS IND.	0.365	-0.396	0.133	
TOT. CORRECT	0.341			

TEST SAMPLE CLASSIFICATION PROBABILITY TABLE

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	0.523	0.000	0.477	1.000
2	0.488	0.000	0.512	1.000
3	0.421	0.000	0.579	1.000

LEARNING SAMPLE CLASSIFICATION TABLE

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	3122.00	0.00	2039.00	5161.00
2	6305.00	0.00	6072.00	12377.00
3	5109.00	0.00	7834.00	12943.00
PRED. TOT.	14536.00	0.00	15945.00	30481.00
CORRECT	0.605	0.000	0.605	
SUCCESS IND.	0.436	-0.406	0.181	
TOT. CORRECT	0.359			

LEARNING SAMPLE CLASSIFICATION PROBABILITY TABLE

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	0.605	0.000	0.395	1.000
2	0.509	0.000	0.491	1.000
3	0.395	0.000	0.605	1.000

Output A12. Ketepatan Klasifikasi 70% : 30% dengan Indeks Twoing

=====

TEST SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	739.00	593.00	731.00	2063.00
2	1783.00	1534.00	1851.00	5168.00
3	1776.00	1469.00	2587.00	5832.00
PRED. TOT.	4298.00	3596.00	5169.00	13063.00
CORRECT	0.358	0.297	0.444	
SUCCESS IND.	0.200	-0.099	-0.003	
TOT. CORRECT	0.372			

=====

TEST SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	0.358	0.287	0.354	1.000
2	0.345	0.297	0.358	1.000
3	0.305	0.252	0.444	1.000

=====

LEARNING SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	2754.00	996.00	1411.00	5161.00
2	4158.00	3552.00	4667.00	12377.00
3	3326.00	2652.00	6965.00	12943.00
PRED. TOT.	10238.00	7200.00	13043.00	30481.00
CORRECT	0.534	0.287	0.538	
SUCCESS IND.	0.364	-0.119	0.114	
TOT. CORRECT	0.435			

=====

LEARNING SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	0.534	0.193	0.273	1.000
2	0.336	0.287	0.377	1.000
3	0.257	0.205	0.538	1.000

Output A13. Ketepatan Klasifikasi *Cross Validation 10-fold* dengan Indeks Gini

```
=====
```

CROSS VALIDATION CLASSIFICATION TABLE				
=====				
Actual Class	Predicted Class			Actual Total
	1	2	3	
1	3199.00	1598.00	2427.00	7224.00
2	5970.00	4381.00	7194.00	17545.00
3	5088.00	3981.00	9706.00	18775.00
PRED. TOT.	14257.00	9960.00	19327.00	43544.00
CORRECT	0.443	0.250	0.517	
SUCCESS IND.	0.277	-0.153	0.086	
TOT. CORRECT	0.397			

```
-----
```

```
=====
```

CROSS VALIDATION CLASSIFICATION PROBABILITY TABLE				
=====				
Actual Class	Predicted Class			Actual Total
	1	2	3	
1	0.443	0.221	0.336	1.000
2	0.340	0.250	0.410	1.000
3	0.271	0.212	0.517	1.000

```
-----
```

```
=====
```

LEARNING SAMPLE CLASSIFICATION TABLE				
=====				
Actual Class	Predicted Class			Actual Total
	1	2	3	
1	3726.00	1267.00	2231.00	7224.00
2	5654.00	4674.00	7217.00	17545.00
3	4853.00	3429.00	10493.00	18775.00
PRED. TOT.	14233.00	9370.00	19941.00	43544.00
CORRECT	0.516	0.266	0.559	
SUCCESS IND.	0.350	-0.137	0.128	
TOT. CORRECT	0.434			

```
-----
```

```
=====
```

LEARNING SAMPLE CLASSIFICATION PROBABILITY TABLE				
=====				
Actual Class	Predicted Class			Actual Total
	1	2	3	
1	0.516	0.175	0.309	1.000
2	0.322	0.266	0.411	1.000
3	0.258	0.183	0.559	1.000

```
-----
```

Output A14. Ketepatan Klasifikasi *Cross Validation 10-fold* dengan Indeks Twoing

=====

CROSS VALIDATION CLASSIFICATION TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	3272.00	1346.00	2606.00	7224.00
2	6102.00	3848.00	7595.00	17545.00
3	5111.00	3517.00	10147.00	18775.00
PRED. TOT.	14485.00	8711.00	20348.00	43544.00
CORRECT	0.453	0.219	0.540	
SUCCESS IND.	0.287	-0.184	0.109	
TOT. CORRECT	0.397			

=====

CROSS VALIDATION CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	0.453	0.186	0.361	1.000
2	0.348	0.219	0.433	1.000
3	0.272	0.187	0.540	1.000

=====

LEARNING SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	3597.00	1205.00	2422.00	7224.00
2	6250.00	4157.00	7138.00	17545.00
3	4995.00	3687.00	10093.00	18775.00
PRED. TOT.	14842.00	9049.00	19653.00	43544.00
CORRECT	0.498	0.237	0.538	
SUCCESS IND.	0.332	-0.166	0.106	
TOT. CORRECT	0.410			

=====

LEARNING SAMPLE CLASSIFICATION PROBABILITY TABLE

=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	0.498	0.167	0.335	1.000
2	0.356	0.237	0.407	1.000
3	0.266	0.196	0.538	1.000

Lampiran B. Pohon Klasifikasi Maksimal

Output B1. Informasi Pembentukan Pohon Klasifikasi dengan *Cross Validation 10-fold*

Tree Sequence				
Tree Number	Terminal Nodes	Cross-Validated Relative Cost	Resubstitution Relative Cost	Complexity
1	7842	0.934 ± 0.004	0.296	-1.000
2	7370	0.933 ± 0.004	0.299	1.00E-005
3	5861	0.928 ± 0.004	0.339	2.01E-005
4	4813	0.926 ± 0.004	0.382	3.02E-005
5	3524	0.923 ± 0.004	0.454	4.04E-005
6	2601	0.921 ± 0.004	0.519	5.04E-005
7	1843	0.920 ± 0.004	0.583	6.04E-005
8	1467	0.916 ± 0.004	0.620	7.05E-005
9	1063	0.911 ± 0.004	0.666	8.06E-005
10	852	0.908 ± 0.004	0.694	9.07E-005
11	656	0.907 ± 0.004	0.722	0.000101
12	530	0.906 ± 0.004	0.742	0.000111
13	434	0.904 ± 0.004	0.759	0.000122
14	366	0.901 ± 0.004	0.772	0.000132
15	315	0.901 ± 0.004	0.783	0.000143
16	264	0.900 ± 0.004	0.794	0.000154
17	220	0.898 ± 0.004	0.805	0.000164
18	201	0.897 ± 0.004	0.809	0.000176
19	185	0.896 ± 0.004	0.814	0.000188
20	167	0.897 ± 0.004	0.819	0.000198
21	158	0.896 ± 0.004	0.822	0.000209
22	150	0.894 ± 0.004	0.824	0.000220
23	145	0.896 ± 0.004	0.826	0.000230
24	141	0.894 ± 0.004	0.827	0.000242
25	116	0.895 ± 0.004	0.837	0.000252
26	110	0.894 ± 0.004	0.839	0.000264
27	86	0.894 ± 0.004	0.849	0.000274
28	76	0.894 ± 0.004	0.853	0.000290
29	69	0.894 ± 0.004	0.856	0.000303
30	58	0.894 ± 0.004	0.861	0.000322
31**	53	0.894 ± 0.004	0.864	0.000364
32	45	0.895 ± 0.004	0.868	0.000384
33	40	0.895 ± 0.004	0.871	0.000417
34	39	0.895 ± 0.004	0.872	0.000437
35	37	0.894 ± 0.004	0.873	0.000450
36	30	0.896 ± 0.004	0.878	0.000478
37	29	0.897 ± 0.004	0.879	0.000507
38	23	0.898 ± 0.004	0.884	0.000524
39	17	0.898 ± 0.004	0.889	0.000572
40	15	0.900 ± 0.004	0.891	0.000674
41	10	0.901 ± 0.004	0.896	0.000686
42	9	0.898 ± 0.004	0.897	0.000818
43	8	0.903 ± 0.004	0.899	0.001
44	5	0.907 ± 0.004	0.906	0.001
45	4	0.916 ± 0.003	0.913	0.005
46	3	0.926 ± 0.003	0.922	0.006
47	2	0.932 ± 0.003	0.932	0.007
48	1	1.000 ± 8.82E-005	1.000	0.046

* Minimum Cost

** Optimal

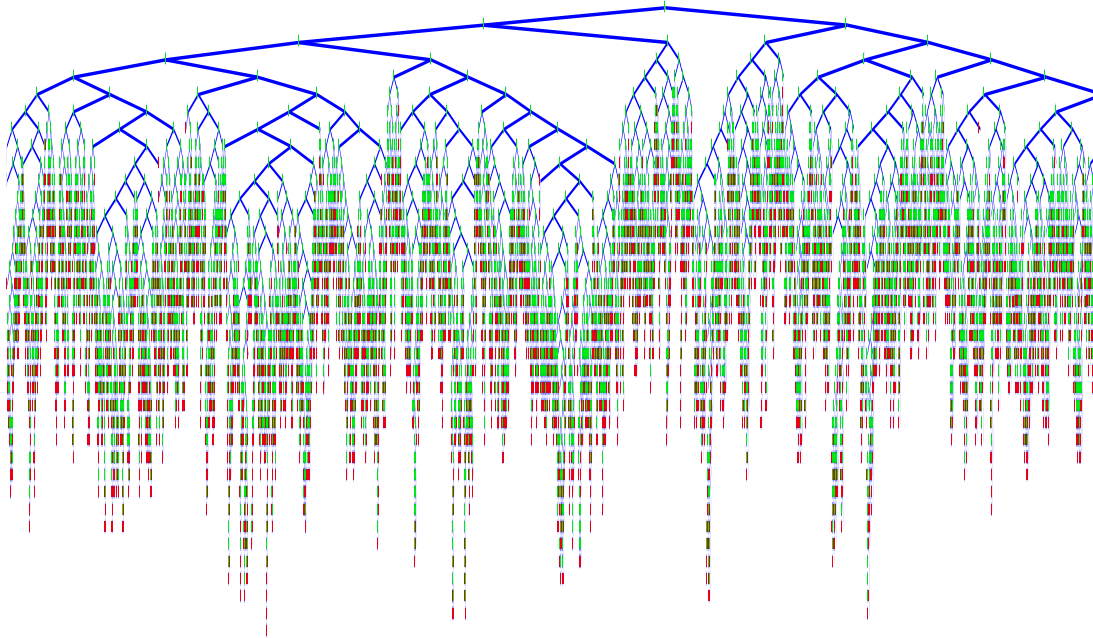
Output B2. Informasi Pemangkasan Pohon Klasifikasi dengan *Cross Validation 10-fold*

```
=====
TREE SEQUENCE
=====
```

Dependent variable: Y

Terminal Tree Nodes		Cross-Validated Relative Cost	Resubstitution Relative Cost	Complexity Parameter

1	7842	0.934 +/- 0.004	0.296	0.000
31**	53	0.894 +/- 0.004	0.864	.364288E-03
39	17	0.898 +/- 0.004	0.889	.571584E-03
40	15	0.900 +/- 0.004	0.891	.673796E-03
41	10	0.901 +/- 0.004	0.896	.685886E-03
42	9	0.898 +/- 0.004	0.897	.817934E-03
43	8	0.903 +/- 0.004	0.899	0.001
44	5	0.907 +/- 0.004	0.906	0.001
45	4	0.916 +/- 0.003	0.913	0.005
46	3	0.926 +/- 0.003	0.922	0.006
47	2	0.932 +/- 0.003	0.932	0.007
48	1	1.000 +/- .882174E-04	1.000	0.046
** Optimal				

Output B3. Gambar Topologi Pohon Klasifikasi Maksimal

Output B4. Ketepatan Klasifikasi Data *Learning* dan Data *Testing* Pohon Klasifikasi Maksimal

Misclassification for Learn Data

Class	N Cases	N Mis-classed	Pct Error	Cost
1	7224	518	7.17	0.07
2	17545	4228	24.10	0.24
3	18775	5242	27.92	0.28

			Predicted Class		
Actual Class	Total Cases	Percent Correct	Class 1 N=11262	Class 2 N=16609	Class 3 N=15673
1	7,224	92.829	6,706	295	223
2	17,545	75.902	2,311	13,317	1,917
3	18,775	72.080	2,245	2,997	13,533

Misclassification for Test Data

Class	N Cases	N Mis-classed	Pct Error	Cost
1	7224	4863	67.32	0.67
2	17545	10764	61.35	0.61
3	18775	10905	58.08	0.58

			Predicted Class		
Actual Class	Total Cases	Percent Correct	Class 1 N=11240	Class 2 N=15998	Class 3 N=16306
1	7,224	32.683	2,361	2,604	2,259
2	17,545	38.649	4,587	6,781	6,177
3	18,775	41.917	4,292	6,613	7,870

Lampiran C. Pohon Klasifikasi Optimal

Output C1. Ketepatan Klasifikasi Data *Learning* dan Data *Testing* Pohon Klasifikasi Optimal

Misclassification for Learn Data

Class	N Cases	N Mis-classed	Pct Error	Cost
1	7224	3627	50.21	0.50
2	17545	13388	76.31	0.76
3	18775	8682	46.24	0.46

			Predicted Class		
Actual Class	Total Cases	Percent Correct	Class 1 N=14842	Class 2 N=9049	Class 3 N=19653
1	7,224	49.792	3,597	1,205	2,422
2	17,545	23.693	6,250	4,157	7,138
3	18,775	53.758	4,995	3,687	10,093

Misclassification for Test Data

Class	N Cases	N Mis-classed	Pct Error	Cost
1	7224	3952	54.71	0.55
2	17545	13697	78.07	0.78
3	18775	8628	45.95	0.46

			Predicted Class		
Actual Class	Total Cases	Percent Correct	Class 1 N=14485	Class 2 N=8711	Class 3 N=20348
1	7,224	45.293	3,272	1,346	2,606
2	17,545	21.932	6,102	3,848	7,595
3	18,775	54.045	5,111	3,517	10,147

Output C2. Informasi Pemilahan Simpul Pohon Klasifikasi Optimal

=====
 NODE INFORMATION
 =====

```

*****
*                               *
*           Node 1: PENGHASI    *
*           N: 43544            *
*                               *
*****

*****                               *****
*                               *           *
*           Node 2              *           *           Node 26      *
*           N: 24745            *           *           N: 18799      *
*                               *           *           *           *
*****                               *****

Node 1 was split on PENGHASI
A case goes left if PENGHASI <= 282000.000
Improvement = 0.010                      Complexity Threshold = 0.046

Node      Cases      Wgt Counts      Cost      Class
1         43544      43544.00      0.667 1
2         24745      24745.00      0.622 3
26        18799      18799.00      0.619 1

Weighted Counts
Class      Top      Left      Right
1          7224.00   3522.00   3702.00
2          17545.00  9502.00   8043.00
3          18775.00  11721.00  7054.00

Within Node Probabilities
Class      Top      Left      Right
1          0.333    0.295    0.381
2          0.333    0.328    0.340
3          0.333    0.378    0.279

Surrogate      Split      Assoc.      Improve.
1 X16          s 1      0.260      0.003
2 KEPEMILI    s 159000.000 0.247      0.006
3 X11         s 1,3,4,5 0.090      .629643E-05
4 KONSUMSI    s 1.500    0.034      .320157E-03
5 V20_A       s 1.500    0.021      .192438E-03

Competitor      Split      Improve.
1 KEPEMILI      25000.000
2 X15           2,3,4      0.008
3 X16           1,4,6      0.005
4 X5            1,3      0.004
5 V20_A         0.500      0.002

```

```

*****
*                               *
*           Node 2: X5          *
*           N: 24745            *
*                               *
*****

*****                               *****
*                               *           *
*           Node 3              *           *           Node 22      *
*           N: 21984            *           *           N: 2761      *
*                               *           *           *           *
*****                               *****

```

```

Node 2 was split on X5
A case goes left if X5 = (1,3)
Improvement = 0.003                      Complexity Threshold = 0.006

Node      Cases      Wgt Counts      Cost      Class
2         24745      24745.00      0.622 3
3         21984      21984.00      0.608 3
22        2761      2761.00      0.633 1

Weighted Counts
Class      Top      Left      Right
1          3522.00   3003.00   519.00
2          9502.00   8267.00   1235.00
3          11721.00  10714.00  1007.00

```

Within Node Probabilities			
Class	Top	Left	Right
1	0.295	0.285	0.367
2	0.328	0.323	0.359
3	0.378	0.392	0.274

Competitor	Split	Improve.
1 KEPEMILI	25000.000	0.003
2 LUASKAVL	20.500	0.002
3 X8	2,4,5	0.001
4 X15	2,3,4	0.001
5 X10	5,6	0.001

```

*****
*                               *
*           Node 3: KEPEMILI   *
*           N: 21984           *
*                               *
*****
*****
*                               *
*           Node 4             *
*           N: 13610           *
*                               *
*****
*                               *
*           Node 8             *
*           N: 8374            *
*                               *
*****

```

Node 3 was split on KEPEMILI
A case goes left if KEPEMILI <= 25000.000
Improvement = 0.002 Complexity Threshold = 0.001

Node	Cases	Wgt Counts	Cost	Class
3	21984	21984.00	0.608	3
4	13610	13610.00	0.580	3
8	8374	8374.00	0.652	3

Weighted Counts			
Class	Top	Left	Right
1	3003.00	1659.00	1344.00
2	8267.00	4979.00	3288.00
3	10714.00	6972.00	3742.00

Within Node Probabilities			
Class	Top	Left	Right
1	0.285	0.260	0.325
2	0.323	0.321	0.327
3	0.392	0.420	0.348

Surrogate	Split	Assoc.	Improve.
1 PENGHASI s	102500.000	0.080	.325493E-03

Competitor	Split	Improve.
1 LUASKAVL	20.500	0.001
2 X7	4	0.001
3 LUASLANT	20.500	0.001
4 X8	1,2,4,5	0.001
5 X16	1,5,6	0.001

```

*****
*                               *
*           Node 4: X8         *
*           N: 13610           *
*                               *
*****
*****
*                               *
*           Node 5             *
*           N: 6310            *
*                               *
*****
=====
*                               *
*           Node 5             *
*           N: 6310            *
*                               *
=====

```

Node 4 was split on X8
A case goes left if X8 = (1,5)
Improvement = .878820E-03 Complexity Threshold = .514349E-03

Node	Cases	Wgt Counts	Cost	Class
4	13610	13610.00	0.580	3
5	6310	6310.00	0.605	3
-5	7300	7300.00	0.558	3

```

                                Weighted Counts
Class      Top      Left      Right
1          1659.00    873.00    786.00
2          4979.00    2330.00   2649.00
3          6972.00    3107.00   3865.00

                                Within Node Probabilities
Class      Top      Left      Right
1          0.260     0.288     0.234
2          0.321     0.317     0.324
3          0.420     0.395     0.442

Surrogate      Split      Assoc.      Improve.
1 X7           s 1,2,3    0.377    .608491E-03
2 X11          s 1,2,3,4  0.131    .286906E-03
3 X6           s 1,2,3    0.117    .125473E-03
4 X9           s 1       0.113    .140094E-03
5 X5           s 1       0.085    .299499E-03

Competitor      Split      Improve.
1 X1            1,4,5,6    .710025E-03
2 X16           1,5,6    .619957E-03
3 X7            4         .608491E-03
4 LUASLANT      21.500    .455157E-03
5 X10           3,6     .423071E-03

*****
*                      Node 5: X5                      *
*                      N: 6310                          *
*****

*****
*                      *      =      Terminal Node 4      =
*                      *      =      N: 4574                =
*                      *      =
*****

Node 5 was split on X5
A case goes left if X5 = (3)
Improvement = .644074E-03      Complexity Threshold = .771530E-03

Node      Cases      Wgt Counts      Cost      Class
5         6310         6310.00         0.605 3
6         1736         1736.00         0.659 3
-4        4574         4574.00         0.584 3

                                Weighted Counts
Class      Top      Left      Right
1          873.00    280.00    593.00
2          2330.00    695.00    1635.00
3          3107.00    761.00    2346.00

                                Within Node Probabilities
Class      Top      Left      Right
1          0.288     0.326     0.273
2          0.317     0.333     0.310
3          0.395     0.341     0.416

Surrogate      Split      Assoc.      Improve.
1 X6           s 4,5,6    0.423    .241459E-03
2 X7           s 4       0.075    .109225E-03
3 X8           s 5       0.030    .851090E-04

Competitor      Split      Improve.
1 X1            4,5,6    .588304E-03
2 X6            3,4,5,6 .567752E-03
3 X16           1,2,4,6 .320199E-03
4 X9            2,4     .312754E-03
5 KONSUMSI      3.500    .311226E-03

```



```

*****
*                               *
*           Node 6: X11         *
*           N: 1736             *
*                               *
*****

*****
*                               *
*           Node 7             *
*           N: 1351            *
*                               *
*****

Node 6 was split on X11
A case goes left if X11 = (1,3,5)
Improvement = .395758E-03      Complexity Threshold = 0.002

Node      Cases      Wgt Counts      Cost Class
6         1736       1736.00         0.659 3
7         1351       1351.00         0.631 3
-3        385        385.00         0.585 1

Weighted Counts
Class      Top      Left      Right
1          280.00    195.00    85.00
2          695.00    529.00    166.00
3          761.00    627.00    134.00

Within Node Probabilities
Class      Top      Left      Right
1          0.326    0.298    0.415
2          0.333    0.333    0.334
3          0.341    0.369    0.252

Surrogate      Split      Assoc.      Improve.
1 X16          s 1,2,5,6    0.042      .238635E-03
2 X1           s 1,4,5,6    0.026      .263113E-03

Competitor      Split      Improve.
1 LUASLANT      27.500      .379015E-03
2 X1            2,3,4,6      .351510E-03
3 X16           1,5,6       .302059E-03
4 LUASKAVL      322.000      .287795E-03
5 X10           3,6         .285963E-03

*****
*                               *
*           Node 7: LUASLANT   *
*           N: 1351            *
*                               *
*****

=====
=          Terminal Node 1      =          Terminal Node 2      =
=          N: 567               =          N: 784               =
=====

Node 7 was split on LUASLANT
A case goes left if LUASLANT <= 27.500
Improvement = .344438E-03      Complexity Threshold = .427246E-03

Node      Cases      Wgt Counts      Cost Class
7         1351       1351.00         0.631 3
-1        567        567.00         0.567 3
-2        784        784.00         0.651 2

Weighted Counts
Class      Top      Left      Right
1          195.00    68.00    127.00
2          529.00    200.00   329.00
3          627.00    299.00   328.00

Within Node Probabilities
Class      Top      Left      Right
1          0.298    0.256    0.327
2          0.333    0.310    0.349
3          0.369    0.433    0.325

Surrogate      Split      Assoc.      Improve.
1 LUASKAVL      s 42.500      0.416      .103624E-04
2 X1           s 2,3,5      0.020      .700420E-04

```

3	KONSUMSI	r	2.500	0.011	.445926E-04
4	PENGHASI	r	205000.000	0.011	.752807E-04
5	X16	s	3,5,6	0.010	.819875E-04

Competitor			Split	Improve.	
1	X1		1,5,6		.295150E-03
2	LUASKAVL		322.500		.223665E-03
3	X16		3,4,6		.147450E-03
4	KONSUMSI		3.500		.143127E-03
5	X7		1,3,4		.141970E-03

```

*****
*                               *
*           Node 8: LUASLANT    *
*           N: 8374             *
*                               *
*****

```

```

=====
=           Terminal Node 6      = *           Node 9          *
=           N: 475              = *           N: 7899         *
=====

```

Node 8 was split on LUASLANT

A case goes left if LUASLANT <= 4.000

Improvement = 0.001 Complexity Threshold = 0.002

Node	Cases	Wgt	Counts	Cost	Class
8	8374		8374.00	0.652	3
-6	475		475.00	0.494	3
9	7899		7899.00	0.661	3

Weighted Counts

Class	Top	Left	Right
1	1344.00	42.00	1302.00
2	3288.00	153.00	3135.00
3	3742.00	280.00	3462.00

Within Node Probabilities

Class	Top	Left	Right
1	0.325	0.197	0.332
2	0.327	0.296	0.329
3	0.348	0.506	0.339

Surrogate		Split	Assoc.	Improve.
1	LUASKAVL s	6.500	0.096	.774412E-03
Competitor		Split		Improve.
1	LUASKAVL	17.500		.997742E-03
2	X15	1,4		.709841E-03
3	X8	1,3,5		.680615E-03
4	KEPEMILI	77000.000		.528514E-03
5	X6	1,3,5,6		.499327E-03

```

*****
*                               *
*           Node 9: X8          *
*           N: 7899             *
*                               *
*****

```

```

=====
*           Node 10             * *           Node 15          *
*           N: 2791             * *           N: 5108         *
=====

```

Node 9 was split on X8

A case goes left if X8 = (2,4,5)

Improvement = .814403E-03 Complexity Threshold = 0.004

Node	Cases	Wgt	Counts	Cost	Class
9	7899		7899.00	0.661	3
10	2791		2791.00	0.637	1
15	5108		5108.00	0.635	3

Weighted Counts

Class	Top	Left	Right
1	1302.00	516.00	786.00

```

2          3135.00      1184.00      1951.00
3          3462.00      1091.00      2371.00
          Within Node Probabilities
Class      Top      Left      Right
1          0.332      0.363      0.314
2          0.329      0.343      0.321
3          0.339      0.295      0.365

          Surrogate      Split      Assoc.      Improve.
1 X7      s 2          0.046      .195895E-03
          Competitor      Split      Improve.
1 X15      1,4          .696176E-03
2 KEPEMILI      77000.000      .532159E-03
3 X6      1,3,6      .527124E-03
4 V21_A      2.500      .427229E-03
5 X7      4          .418419E-03

          *****
          *          Node 10: X7          *
          *          N: 2791          *
          *****

          *****
          *          Node 11          *          *          Node 14          *
          *          N: 1065          *          *          N: 1726          *
          *****

Node 10 was split on X7
A case goes left if X7 = (4)
Improvement = .620908E-03      Complexity Threshold = 0.001
Node      Cases      Wgt Counts      Cost      Class
10      2791      2791.00      0.637 1
11      1065      1065.00      0.644 2
14      1726      1726.00      0.601 1

          Weighted Counts
Class      Top      Left      Right
1          516.00      154.00      362.00
2          1184.00      446.00      738.00
3          1091.00      465.00      626.00
          Within Node Probabilities
Class      Top      Left      Right
1          0.363      0.298      0.399
2          0.343      0.356      0.335
3          0.295      0.346      0.266

          Surrogate      Split      Assoc.      Improve.
1 X8      s 2,5          0.151      .284216E-04
2 LUASLANT s      22.500      0.026      .232746E-03
          Competitor      Split      Improve.
1 X15      1,4          .613890E-03
2 X6      1,3,4,5,6      .509651E-03
3 KEPEMILI      77500.000      .359220E-03
4 X10      5,6      .279376E-03
5 LUASLANT      29.500      .272821E-03

          *****
          *          Node 11: KEPEMILI          *
          *          N: 1065          *
          *****

          *****
          *          Terminal Node 7          *          *          Node 12          *
          *          N: 82          *          *          N: 983          *
          *****

Node 11 was split on KEPEMILI
A case goes left if KEPEMILI <= 77500.000
Improvement = .640221E-03      Complexity Threshold = .675886E-03
Node      Cases      Wgt Counts      Cost      Class
11      1065      1065.00      0.644 2

```

```

-7      82      82.00      0.449 1
12     983      983.00      0.634 2

      Weighted Counts
Class      Top      Left      Right
1         154.00     27.00     127.00
2         446.00     30.00     416.00
3         465.00     25.00     440.00

      Within Node Probabilities
Class      Top      Left      Right
1         0.298     0.551     0.272
2         0.356     0.252     0.366
3         0.346     0.196     0.362

      Competitor      Split      Improve.
1 X6      3,4,6      75.500      .311156E-03
2 LUASLANT      37.000      .252376E-03
3 LUASKAVL      37.000      .245059E-03
4 X15      1,4      .169172E-03
5 X5      3      .157254E-03

*****
*      Node 12: KEPEMILI      *
*      N: 983      *
*****

*****
*      Node 13      *      =      Terminal Node 10      =
*      N: 936      *      =      N: 47      =
*****

Node 12 was split on KEPEMILI
A case goes left if KEPEMILI <= 3825000.000
Improvement = .315632E-03      Complexity Threshold = .440466E-03

Node      Cases      Wgt Counts      Cost      Class
12      983      983.00      0.634 2
13      936      936.00      0.628 2
-10     47      47.00      0.485 1

      Weighted Counts
Class      Top      Left      Right
1         127.00     113.00     14.00
2         416.00     398.00     18.00
3         440.00     425.00     15.00

      Within Node Probabilities
Class      Top      Left      Right
1         0.272     0.257     0.515
2         0.366     0.372     0.273
3         0.362     0.371     0.212

      Competitor      Split      Improve.
1 LUASLANT      75.500      .266011E-03
2 LUASKAVL      39.000      .203584E-03
3 X4      2,3,4,5,6      .122325E-03
4 X15      1,4      .998721E-04
5 X5      3      .964259E-04

*****
*      Node 13: KEPEMILI      *
*      N: 936      *
*****

=====
=      Terminal Node 8      =      Terminal Node 9      =
=      N: 789      =      N: 147      =
=====

Node 13 was split on KEPEMILI
A case goes left if KEPEMILI <= 825000.000
Improvement = .275448E-03      Complexity Threshold = .576914E-03

Node      Cases      Wgt Counts      Cost      Class

```

	13	936	936.00	0.628	2
	-8	789	789.00	0.619	2
	-9	147	147.00	0.488	3

Weighted Counts

Class	Top	Left	Right
1	113.00	102.00	11.00
2	398.00	348.00	50.00
3	425.00	339.00	86.00

Within Node Probabilities

Class	Top	Left	Right
1	0.257	0.271	0.170
2	0.372	0.381	0.318
3	0.371	0.347	0.512

Surrogate	Split	Assoc.	Improve.
1 X11	s 1,2,3,5	0.033	.202400E-04
2 LUASLANT	s 147.500	0.012	.291943E-04
Competitor	Split		Improve.
1 LUASLANT	75.500		.274378E-03
2 LUASKAVL	37.000		.201199E-03
3 X4	1,4,6		.136833E-03
4 X1	2,3,4,5,6		.121906E-03
5 X11	1,4,5		.994674E-04

* Node 14: X15 *
* N: 1726 *

=====		=====	
=	Terminal Node 11	=	Terminal Node 12
=	N: 341	=	N: 1385
=====		=====	

Node 14 was split on X15
A case goes left if X15 = (2,3,4)
Improvement = .471980E-03 Complexity Threshold = .681117E-03

Node	Cases	Wgt Counts	Cost	Class
14	1726	1726.00	0.601	1
-11	341	341.00	0.624	2
-12	1385	1385.00	0.576	1

Weighted Counts

Class	Top	Left	Right
1	362.00	47.00	315.00
2	738.00	150.00	588.00
3	626.00	144.00	482.00

Within Node Probabilities

Class	Top	Left	Right
1	0.399	0.286	0.424
2	0.335	0.376	0.326
3	0.266	0.337	0.250

Competitor	Split	Improve.
1 X10	5,6	.277144E-03
2 KONSUMSI	0.500	.270627E-03
3 X6	2,4,6	.229541E-03
4 LUASKAVL	154.500	.225475E-03
5 V20_A	0.500	.200502E-03

* Node 15: V21_A *
* N: 5108 *

*****		*****	
*	Node 16	*	Node 17
*	N: 413	*	N: 4695
*****		*****	

Node 15 was split on V21_A

A case goes left if V21_A <= 2.500

Improvement = .463288E-03 Complexity Threshold = .807934E-03

Node	Cases	Wgt	Counts	Cost	Class
15	5108		5108.00	0.635	3
16	413		413.00	0.583	1
17	4695		4695.00	0.633	3

Weighted Counts

Class	Top	Left	Right
1	786.00	91.00	695.00
2	1951.00	131.00	1820.00
3	2371.00	191.00	2180.00

Within Node Probabilities

Class	Top	Left	Right
1	0.314	0.417	0.304
2	0.321	0.247	0.328
3	0.365	0.336	0.367

Competitor	Split	Improve.
1 X10	6	.300866E-03
2 X6	1,6	.298734E-03
3 LUASKAVL	47.500	.287795E-03
4 X11	2,5	.240500E-03
5 X7	1,4	.225776E-03

 * Node 16: LUASKAVL *
 * N: 413 *

Terminal Node 13	Terminal Node 14
N: 88	N: 325

Node 16 was split on LUASKAVL

A case goes left if LUASKAVL <= 34.000

Improvement = 0.001 Complexity Threshold = .515778E-03

Node	Cases	Wgt	Counts	Cost	Class
16	413		413.00	0.583	1
-13	88		88.00	0.311	1
-14	325		325.00	0.615	3

Weighted Counts

Class	Top	Left	Right
1	91.00	41.00	50.00
2	131.00	15.00	116.00
3	191.00	32.00	159.00

Within Node Probabilities

Class	Top	Left	Right
1	0.417	0.689	0.315
2	0.247	0.104	0.301
3	0.336	0.207	0.385

Surrogate	Split	Assoc.	Improve.
1 LUASLANT	29.000	0.480	.849651E-03
2 KEPEMILI	77500.000	0.313	.723120E-03
3 X6	s 5	0.041	.766491E-04
4 X4	s 3,6	0.017	.631057E-04

Competitor	Split	Improve.
1 LUASLANT	29.000	.849651E-03
2 KEPEMILI	90000.000	.818592E-03
3 X6	1,3,4,6	.348511E-03
4 X8	3	.328018E-03
5 X11	1,5	.217779E-03

```

*****
*                               *
*           Node 17: X10        *
*           N: 4695             *
*                               *
*****
*****
*                               *
*           Node 18            *
*           N: 1131            *
*                               *
*****
*                               *
*           Node 21            *
*           N: 3564            *
*                               *
*****

```

Node 17 was split on X10

A case goes left if X10 = (1,3,6)

Improvement = .321631E-03 Complexity Threshold = .663796E-03

Node	Cases	Wgt Counts	Cost	Class
17	4695	4695.00	0.633	3
18	1131	1131.00	0.648	1
21	3564	3564.00	0.621	3

Weighted Counts

Class	Top	Left	Right
1	695.00	201.00	494.00
2	1820.00	436.00	1384.00
3	2180.00	494.00	1686.00

Within Node Probabilities

Class	Top	Left	Right
1	0.304	0.352	0.288
2	0.328	0.315	0.333
3	0.367	0.333	0.379

Competitor	Split	Improve.
1 X7	1,4	.297079E-03
2 X15	2,3,4	.274304E-03
3 X1	3,4,6	.221174E-03
4 X16	1,5,6	.220776E-03
5 X11	2,5	.208957E-03

```

*****
*                               *
*           Node 18: X6        *
*           N: 1131            *
*                               *
*****
*****
*                               *
*           Node 19            *
*           N: 873             *
*                               *
*****
*****
*                               *
*           Node 19            *
*           N: 873             *
*                               *
*****
=====
*                               *
*           Node 19            *
*           N: 873             *
*                               *
*****
=====
*                               *
*           Terminal Node 18    *
*           N: 258             *
*                               *
=====

```

Node 18 was split on X6

A case goes left if X6 = (1,3,5,6)

Improvement = .322158E-03 Complexity Threshold = .467661E-03

Node	Cases	Wgt Counts	Cost	Class
18	1131	1131.00	0.648	1
19	873	873.00	0.658	3
-18	258	258.00	0.551	1

Weighted Counts

Class	Top	Left	Right
1	201.00	138.00	63.00
2	436.00	353.00	83.00
3	494.00	382.00	112.00

Within Node Probabilities

Class	Top	Left	Right
1	0.352	0.321	0.449
2	0.315	0.338	0.244
3	0.333	0.342	0.307

Surrogate	Split	Assoc.	Improve.
1 X1	s 1,3,4,5,6	0.016	.600037E-04
Competitor	Split		Improve.
1 X16	3,5,6		.277112E-03

```

2 X15          2,3,4          .255493E-03
3 X10          6             .249401E-03
4 LUASLANT     26.000        .239446E-03
5 PENGHASI     177500.000    .207320E-03

```

```

*****
*                   Node 19: X10          *
*                   N: 873                *
*****

```

```

=====
=      Terminal Node 15      =      *      Node 20      *
=      N: 21                 =      *      N: 852        *
=====
*****

```

Node 19 was split on X10

A case goes left if X10 = (6)

Improvement = .282300E-03 Complexity Threshold = .494289E-03

Node	Cases	Wgt	Counts	Cost	Class
19	873		873.00	0.658	3
-15	21		21.00	0.353	1
20	852		852.00	0.648	3

Weighted Counts

Class	Top	Left	Right
1	138.00	9.00	129.00
2	353.00	11.00	342.00
3	382.00	1.00	381.00

Within Node Probabilities

Class	Top	Left	Right
1	0.321	0.647	0.310
2	0.338	0.326	0.338
3	0.342	0.028	0.352

Surrogate	Split	Assoc.	Improve.
1 LUASKAVL r	592.500	0.042	.392859E-04

Competitor	Split	Improve.
1 PENGHASI	177500.000	.207949E-03
2 X15	2,3,4	.199433E-03
3 X16	3,5,6	.197597E-03
4 KEPEMILI	260000.000	.182641E-03
5 LUASLANT	26.000	.167012E-03

```

*****
*                   Node 20: X15          *
*                   N: 852                *
*****

```

```

=====
=      Terminal Node 16      =      =      Terminal Node 17      =
=      N: 209                 =      =      N: 643                =
=====
=====

```

Node 20 was split on X15

A case goes left if X15 = (2,3,4)

Improvement = .199663E-03 Complexity Threshold = .591051E-03

Node	Cases	Wgt	Counts	Cost	Class
20	852		852.00	0.648	3
-16	209		209.00	0.571	2
-17	643		643.00	0.632	3

Weighted Counts

Class	Top	Left	Right
1	129.00	27.00	102.00
2	342.00	104.00	238.00
3	381.00	78.00	303.00

Within Node Probabilities

Class	Top	Left	Right
1	0.310	0.270	0.322
2	0.338	0.429	0.310
3	0.352	0.301	0.368

Surrogate	Split	Assoc.	Improve.
1 LUASKAVL r	456.000	0.028	.103591E-04
2 PENGHASI r	255000.000	0.012	.138369E-03
Competitor	Split		Improve.
1 X16	3,5,6		.197200E-03
2 KEPEMILI	260000.000		.174961E-03
3 LUASLANT	28.500		.169070E-03
4 X7	1,4		.150448E-03
5 KONSUMSI	3.500		.149771E-03

```

*****
*           Node 21: LUASKAVL           *
*           N: 3564                     *
*****

```

```

=====
=           Terminal Node 19           =           Terminal Node 20           =
=           N: 669                    =           N: 2895                    =
=====

```

Node 21 was split on LUASKAVL

A case goes left if LUASKAVL <= 46.500

Improvement = .285449E-03 Complexity Threshold = .823470E-03

Node	Cases	Wgt Counts	Cost Class
21	3564	3564.00	0.621 3
-19	669	669.00	0.604 2
-20	2895	2895.00	0.612 3

Weighted Counts

Class	Top	Left	Right
1	494.00	84.00	410.00
2	1384.00	305.00	1079.00
3	1686.00	280.00	1406.00

Within Node Probabilities

Class	Top	Left	Right
1	0.288	0.265	0.294
2	0.333	0.396	0.318
3	0.379	0.340	0.388

Surrogate	Split	Assoc.	Improve.
1 LUASLANT s	26.500	0.041	.547394E-04
Competitor	Split		Improve.
1 X15	1,4		.285425E-03
2 X4	1,2,4,6		.209384E-03
3 X1	1,5,6		.175724E-03
4 X6	2,3,4,5,6		.170484E-03
5 X16	1,4,5,6		.167834E-03

```

*****
*           Node 22: X8                 *
*           N: 2761                     *
*****

```

```

=====
=           Terminal Node 21           =           Node 23           *
=           N: 1961                    =           N: 800           *
=====

```

Node 22 was split on X8

A case goes left if X8 = (1,2,3,5)

Improvement = 0.003 Complexity Threshold = 0.005

Node	Cases	Wgt Counts	Cost Class
22	2761	2761.00	0.633 1
-21	1961	1961.00	0.607 2
23	800	800.00	0.473 1

Weighted Counts

Class	Top	Left	Right
1	519.00	272.00	247.00

2	1235.00	903.00	332.00
3	1007.00	786.00	221.00

Within Node Probabilities

Class	Top	Left	Right
1	0.367	0.287	0.527
2	0.359	0.393	0.292
3	0.274	0.320	0.181

Surrogate	Split	Assoc.	Improve.
1 LUASLANT r	4.000	0.109	0.003
2 X9 s 1,2		0.088	0.003
Competitor	Split		Improve.
1 X9	3,4		0.003
2 LUASLANT	8.500		0.003
3 V20_A	0.500		0.001
4 PENGHASI	42500.000		0.001
5 LUASKAVL	48.500		0.001

 * Node 23: LUASLANT *
 * N: 800 *

=====	*****
= Terminal Node 22 =	* Node 24 *
= N: 177 =	* N: 623 *
=====	*****

Node 23 was split on LUASLANT

A case goes left if LUASLANT <= 9.500

Improvement = 0.004 Complexity Threshold = .519674E-03

Node	Cases	Wgt Counts	Cost Class
23	800	800.00	0.473 1
-22	177	177.00	0.140 1
24	623	623.00	0.620 2

Weighted Counts

Class	Top	Left	Right
1	247.00	126.00	121.00
2	332.00	35.00	297.00
3	221.00	16.00	205.00

Within Node Probabilities

Class	Top	Left	Right
1	0.527	0.860	0.376
2	0.292	0.098	0.380
3	0.181	0.042	0.245

Surrogate	Split	Assoc.	Improve.
1 LUASKAVL s	48.500	0.640	0.002
2 X9 s 3,4		0.375	0.002
3 X7 s 4		0.347	0.003
4 X10 s 4		0.217	.681706E-03
5 PENGHASI s	10000.000	0.159	0.001
Competitor	Split		Improve.
1 X7	4		0.003
2 LUASKAVL	48.500		0.002
3 X9	4		0.002
4 X11	4,5		0.001
5 PENGHASI	10000.000		0.001

 * Node 24: X10 *
 * N: 623 *

*****	=====
* Node 25 *	= Terminal Node 25 =

```

*           N: 196           *           =           N: 427           =
*****
Node 24 was split on X10
A case goes left if X10 = (1,5,6)
Improvement = .836141E-03      Complexity Threshold = .979957E-03

Node      Cases      Wgt Counts      Cost Class
 24        623        623.00        0.620 2
 25        196        196.00        0.467 1
 -25        427        427.00        0.604 2

      Weighted Counts
Class      Top      Left      Right
 1         121.00    62.00    59.00
 2         297.00    99.00    198.00
 3         205.00    35.00    170.00

      Within Node Probabilities
Class      Top      Left      Right
 1         0.376    0.533    0.286
 2         0.380    0.351    0.396
 3         0.245    0.116    0.318

      Surrogate      Split      Assoc.      Improve.
 1 X1      s 2,4,5      0.046 .183535E-03
 2 LUASKAVL s      15.500 0.046 .158950E-04
 3 KONSUMSI r      6.500 0.038 .232858E-03
 4 LUASLANT s      15.500 0.015 .551705E-04
 5 X4      s 4,5,6      0.013 .703794E-04

      Competitor      Split      Improve.
 1 KEPEMILI      145000.000 .624230E-03
 2 X11      4,5      .540008E-03
 3 X15      2,3,4      .417695E-03
 4 LUASLANT      113.500 .257051E-03
 5 X1      1,2,3,6      .240097E-03

```

```

*****
*           Node 25: KEPEMILI           *
*           N: 196           *
*****
=====
=      Terminal Node 23      =      Terminal Node 24      =
=      N: 101      =      N: 95      =
=====

```

```

Node 25 was split on KEPEMILI
A case goes left if KEPEMILI <= 65000.000
Improvement = .770261E-03      Complexity Threshold = .496640E-03

```

```

Node      Cases      Wgt Counts      Cost Class
 25        196        196.00        0.467 1
 -23        101        101.00        0.470 2
 -24         95         95.00        0.301 1

      Weighted Counts
Class      Top      Left      Right
 1         62.00    16.00    46.00
 2         99.00    65.00    34.00
 3         35.00    20.00    15.00

      Within Node Probabilities
Class      Top      Left      Right
 1         0.533    0.317    0.699
 2         0.351    0.530    0.213
 3         0.116    0.153    0.088

      Surrogate      Split      Assoc.      Improve.
 1 PENGHASI s      10000.000 0.185 .988630E-04
 2 X11      s 1,2,3      0.152 .208583E-03
 3 LUASLANT r      49.500 0.134 .974748E-04
 4 X10      s 6      0.061 .535463E-04
 5 LUASKAVL s      38.000 0.060 .255005E-03

      Competitor      Split      Improve.

```

1	LUASKAVL	67.500		.439194E-03
2	X1	1,2,3,6		.246843E-03
3	X11	1,3,4,5		.208583E-03
4	X9	2,3,4		.202843E-03
5	LUASLANT	113.500		.193478E-03

```

*****
*                               *
*           Node 26: X15         *
*           N: 18799             *
*                               *
*****

```

```

*****
*                               *
*           Node 27             *
*           N: 3263             *
*                               *
*****

```

```

*****
*                               *
*           Node 35             *
*           N: 15536            *
*                               *
*****

```

Node 26 was split on X15

A case goes left if X15 = (2,3,4)

Improvement = 0.004 Complexity Threshold = 0.007

Node	Cases	Wgt	Counts	Cost	Class
26	18799		18799.00	0.619	1
27	3263		3263.00	0.631	3
35	15536		15536.00	0.600	1

Weighted Counts

Class	Top	Left	Right
1	3702.00	432.00	3270.00
2	8043.00	1336.00	6707.00
3	7054.00	1495.00	5559.00

Within Node Probabilities

Class	Top	Left	Right
1	0.381	0.277	0.400
2	0.340	0.353	0.338
3	0.279	0.369	0.262

Competitor	Split	Improve.
1 PENGHASI	377500.000	0.001
2 KEPEMILI	25000.000	0.001
3 KONSUMSI	0.500	.992220E-03
4 X10	5,6	.961415E-03
5 X11	1,4,5	.898062E-03

```

*****
*                               *
*           Node 27: KONSUMSI   *
*           N: 3263             *
*                               *
*****

```

```

*****
*                               *
*           Node 28             *
*           N: 2725             *
*                               *
*****

```

```

*****
*                               *
*           Node 33             *
*           N: 538              *
*                               *
*****

```

Node 27 was split on KONSUMSI

A case goes left if KONSUMSI <= 1.500

Improvement = .538292E-03 Complexity Threshold = .374406E-03

Node	Cases	Wgt	Counts	Cost	Class
27	3263		3263.00	0.631	3
28	2725		2725.00	0.642	3
33	538		538.00	0.567	3

Weighted Counts

Class	Top	Left	Right
1	432.00	390.00	42.00
2	1336.00	1109.00	227.00
3	1495.00	1226.00	269.00

Within Node Probabilities

Class	Top	Left	Right
1	0.277	0.296	0.176
2	0.353	0.346	0.391
3	0.369	0.358	0.433

Surrogate	Split	Assoc.	Improve.
-----------	-------	--------	----------

```

1 V20_A      s      1.500                      0.082 .118862E-03
Competitor      Split      Improve.
1 KEPEMILI      1137500.000                      .520236E-03
2 LUASKAVL      52.500                      .460471E-03
3 X8      2,4,5                      .390841E-03
4 LUASLANT      48.500                      .292866E-03
5 X11      1,2,3,5                      .258969E-03

*****
*                      Node 28: KEPEMILI                      *
*                      N: 2725                      *
*****

*****
*                      Node 29                      *
*                      N: 2159                      *
*****

*****
*                      Node 32                      *
*                      N: 566                      *
*****

Node 28 was split on KEPEMILI
A case goes left if KEPEMILI <= 1137500.000
Improvement = .584547E-03      Complexity Threshold = .434968E-03

Node      Cases      Wgt Counts      Cost      Class
28      2725      2725.00      0.642 3
29      2159      2159.00      0.652 3
32      566      566.00      0.594 2

Weighted Counts
Class      Top      Left      Right
1      390.00      340.00      50.00
2      1109.00      857.00      252.00
3      1226.00      962.00      264.00

Within Node Probabilities
Class      Top      Left      Right
1      0.296      0.320      0.196
2      0.346      0.332      0.406
3      0.358      0.348      0.398

Competitor      Split      Improve.
1 LUASKAVL      52.500      .464658E-03
2 LUASLANT      48.500      .399743E-03
3 X10      5,6      .344827E-03
4 X11      1,2,3,5      .236095E-03
5 X9      1,2,4      .218396E-03

*****
*                      Node 29: X8                      *
*                      N: 2159                      *
*****

*****
*                      Node 30                      *
*                      N: 1471                      *
*****

*****
*                      Node 29                      *
*                      N: 688                      *
*****

Node 29 was split on X8
A case goes left if X8 = (2,3,4,5)
Improvement = .290014E-03      Complexity Threshold = .553526E-03

Node      Cases      Wgt Counts      Cost      Class
29      2159      2159.00      0.652 3
30      1471      1471.00      0.645 2
-29      688      688.00      0.626 1

Weighted Counts
Class      Top      Left      Right
1      340.00      208.00      132.00
2      857.00      612.00      245.00
3      962.00      651.00      311.00

Within Node Probabilities
Class      Top      Left      Right
1      0.320      0.293      0.374
2      0.332      0.355      0.286
3      0.348      0.353      0.339

Surrogate      Split      Assoc.      Improve.
1 X7      s 2,4      0.243      .182741E-04

```

```

2 X6          s 2,3,4,5,6                      0.013 .620654E-04
Competitor    Split
1 LUASKAVL    52.500                          .272097E-03
2 KEPEMILI    925000.000                      .243121E-03
3 X10         1,2,3,4,6                      .229629E-03
4 LUASLANT    48.500                          .220977E-03
5 X1          4,5,6                          .200719E-03

*****
*                      Node 30: X16            *
*                      N: 1471                 *
*****

=====
=      Terminal Node 26      =      *      Node 31      *
=      N: 1291              =      *      N: 180        *
=====
*****

Node 30 was split on X16
A case goes left if X16 = (1,2,4,5,6)
Improvement = .217476E-03      Complexity Threshold = .510987E-03

Node   Cases   Wgt Counts   Cost   Class
30      1471    1471.00     0.645  2
-26     1291    1291.00     0.632  3
31      180     180.00      0.606  2

Weighted Counts
Class      Top      Left      Right
1          208.00   175.00   33.00
2          612.00   524.00   88.00
3          651.00   592.00   59.00

Within Node Probabilities
Class      Top      Left      Right
1          0.293    0.283    0.359
2          0.355    0.349    0.394
3          0.353    0.368    0.247

Competitor    Split
Improve.
1 X1          4,5,6                      .164008E-03
2 LUASKAVL    509.500                  .153164E-03
3 X8          2,5                      .141658E-03
4 V21_A       2.500                  .138710E-03
5 X9          2,3,4                  .137342E-03

```

```

*****
*                      Node 31: X7            *
*                      N: 180                 *
*****

=====
=      Terminal Node 27      =      =      Terminal Node 28      =
=      N: 98                =      =      N: 82                =
=====
=====

Node 31 was split on X7
A case goes left if X7 = (4)
Improvement = .423346E-03      Complexity Threshold = .466898E-03

Node   Cases   Wgt Counts   Cost   Class
31      180     180.00     0.606  2
-27     98      98.00      0.513  1
-28     82      82.00      0.470  2

Weighted Counts
Class      Top      Left      Right
1          33.00   27.00     6.00
2          88.00   41.00    47.00
3          59.00   30.00    29.00

Within Node Probabilities
Class      Top      Left      Right
1          0.359    0.487    0.164

```

2	0.394	0.305	0.530	
3	0.247	0.208	0.306	

	Surrogate	Split		Assoc.	Improve.
1	X8	s 3,5		0.414	.874767E-04
2	X1	s 1,2,4,5,6		0.060	.128677E-04
3	X4	s 2,4,5,6		0.051	.498458E-05
4	LUASKAVL	r 15.500		0.050	.565263E-04
5	X6	s 2,3,4,5,6		0.044	.148951E-03

	Competitor	Split		Improve.
1	X8	2,4,5		.297363E-03
2	X6	1,4,5,6		.251108E-03
3	X1	1,2,3,6		.142680E-03
4	KEPEMILI	425000.000		.136614E-03
5	PENGHASI	325000.000		.120758E-03

```

*****
*           Node 32: KEPEMILI           *
*           N: 566                       *
*****

```

```

=====
=      Terminal Node 30      =      Terminal Node 31      =
=           N: 520          =           N: 46           =
=====

```

Node 32 was split on KEPEMILI

A case goes left if KEPEMILI <= 9005000.000

Improvement = .308088E-03 Complexity Threshold = .413651E-03

Node	Cases	Wgt	Counts	Cost	Class
32	566		566.00	0.594	2
-30	520		520.00	0.578	2
-31	46		46.00	0.319	3

Weighted Counts

Class	Top	Left	Right
1	50.00	48.00	2.00
2	252.00	242.00	10.00
3	264.00	230.00	34.00

Within Node Probabilities

Class	Top	Left	Right
1	0.196	0.203	0.104
2	0.406	0.422	0.214
3	0.398	0.375	0.681

	Competitor	Split		Improve.
1	LUASKAVL	55.500		.248108E-03
2	X10	5,6		.187707E-03
3	PENGHASI	710000.000		.179622E-03
4	LUASLANT	31.000		.164402E-03
5	X1	1,3,5,6		.155108E-03

```

*****
*           Node 33: LUASLANT           *
*           N: 538                       *
*****

```

```

=====
*      Node 34      *      Terminal Node 34      =
*           N: 510  *           N: 28           =
=====

```

Node 33 was split on LUASLANT

A case goes left if LUASLANT <= 97.500

Improvement = .447139E-03 Complexity Threshold = .410204E-03

Node	Cases	Wgt	Counts	Cost	Class
33	538		538.00	0.567	3
34	510		510.00	0.556	3
-34	28		28.00	0.455	1

Weighted Counts

Class	Top	Left	Right
-------	-----	------	-------

1	42.00	33.00	9.00
2	227.00	220.00	7.00
3	269.00	257.00	12.00

Within Node Probabilities

Class	Top	Left	Right
1	0.176	0.148	0.545
2	0.391	0.407	0.175
3	0.433	0.444	0.280

Surrogate	Split	Assoc.	Improve.
1 KEPEMILI s	.134500E+08	0.037	.613794E-04
2 PENGHASI s	2750000.000	0.025	.282224E-04
Competitor	Split		Improve.
1 X8	4,5		.403851E-03
2 X10	1,3,4,5,6		.359980E-03
3 PENGHASI	575000.000		.251746E-03
4 KONSUMSI	2.500		.243973E-03
5 KEPEMILI	405000.000		.237503E-03

```

*****
*                               *
*           Node 34: X8         *
*           N: 510              *
*                               *
*****

```

```

=====
=      Terminal Node 32      =      Terminal Node 33      =
=      N: 143                =      N: 367                =
=====

```

Node 34 was split on X8

A case goes left if X8 = (4,5)

Improvement = .365136E-03 Complexity Threshold = .618175E-03

Node	Cases	Wgt	Counts	Cost	Class
34	510		510.00	0.556	3
-32	143		143.00	0.439	2
-33	367		367.00	0.516	3

Weighted Counts

Class	Top	Left	Right
1	33.00	6.00	27.00
2	220.00	83.00	137.00
3	257.00	54.00	203.00

Within Node Probabilities

Class	Top	Left	Right
1	0.148	0.098	0.167
2	0.407	0.561	0.349
3	0.444	0.341	0.484

Surrogate	Split	Assoc.	Improve.
1 X1 s 2,3		0.013	.324639E-05
2 PENGHASI r	1675000.000	0.013	.882706E-04
Competitor	Split		Improve.
1 KONSUMSI	2.500		.227171E-03
2 X10	1,4,5,6		.217119E-03
3 KEPEMILI	1025000.000		.214276E-03
4 PENGHASI	575000.000		.188536E-03
5 X6	2,3,4,5,6		.140766E-03

```

*****
*                               *
*           Node 35: KEPEMILI  *
*           N: 15536            *
*                               *
*****

```

```

*****
*                               *
*           Node 36             *
*           N: 4710             *
*                               *
*****

```

```

*****
*                               *
*           Node 39             *
*           N: 10826            *
*                               *
*****

```

Node 35 was split on KEPEMILI

A case goes left if KEPEMILI <= 25000.000

Improvement = 0.001 Complexity Threshold = .677964E-03


```

Node      Cases  Wgt Counts      Cost  Class
 35      15536    15536.00    0.600  1
 36       4710     4710.00    0.623  1
 39      10826    10826.00    0.590  1

      Weighted Counts
Class      Top      Left      Right
1         3270.00    915.00    2355.00
2         6707.00    1880.00   4827.00
3         5559.00    1915.00   3644.00

      Within Node Probabilities
Class      Top      Left      Right
1          0.400    0.377    0.410
2          0.338    0.319    0.346
3          0.262    0.304    0.244

      Competitor      Split      Improve.
1 X11                1,4,5      .903473E-03
2 PENGHASI          377500.000   .880904E-03
3 X8                 4,5      .823383E-03
4 KONSUMSI          0.500      .651562E-03
5 X10                5,6      .643979E-03

*****
*                      Node 36: PENGHASI      *
*                      N: 4710                *
*****

*****
*                      Node 37      *      =      Terminal Node 38      =
*                      N: 2369      *      =      N: 2341                =
*****

Node 36 was split on PENGHASI
A case goes left if PENGHASI <= 355000.000
Improvement = .739915E-03      Complexity Threshold = 0.001

Node      Cases  Wgt Counts      Cost  Class
 36       4710     4710.00    0.623  1
 37       2369     2369.00    0.650  1
-38       2341     2341.00    0.596  1

      Weighted Counts
Class      Top      Left      Right
1          915.00    417.00    498.00
2         1880.00    883.00    997.00
3         1915.00    1069.00   846.00

      Within Node Probabilities
Class      Top      Left      Right
1          0.377    0.350    0.404
2          0.319    0.305    0.333
3          0.304    0.345    0.264

      Surrogate      Split      Assoc.      Improve.
1 X16      s 1,6      0.101   .572109E-04
2 V20_A    s          0.056   .238489E-03
3 X11      s 1,3,4,5  0.055   .278386E-03
4 KONSUMSI s          0.046   .201005E-03
5 LUASLANT s          0.037   .157034E-03

      Competitor      Split      Improve.
1 X7         4      .422416E-03
2 X11        1,4,5  .306342E-03
3 X8         1,2,4,5 .287764E-03
4 X1         1,3,4,5,6 .286637E-03
5 V20_A      0.500  .238489E-03

*****
*                      Node 37: X7      *
*                      N: 2369          *
*****

=====
=      Terminal Node 35      =      *      Node 38      *

```

```

=          N: 1023          =          *          N: 1346          *
=====
Node 37 was split on X7
A case goes left if X7 = (4)
Improvement = .494979E-03      Complexity Threshold = 0.002

Node      Cases      Wgt Counts      Cost      Class
 37      2369      2369.00      0.650 1
-35      1023      1023.00      0.599 3
 38      1346      1346.00      0.623 1

      Weighted Counts
Class      Top      Left      Right
1          417.00      156.00      261.00
2          883.00      347.00      536.00
3         1069.00      520.00      549.00

      Within Node Probabilities
Class      Top      Left      Right
1          0.350      0.313      0.377
2          0.305      0.286      0.318
3          0.345      0.401      0.305

      Surrogate      Split      Assoc.      Improve.
1 X8      s 2,3,5      0.647      .220281E-03
2 X6      s 4,5,6      0.114      .654895E-04
3 X5      s 2,3      0.030      .483874E-04
4 X10     s 5      0.017      .155858E-04

      Competitor      Split      Improve.
1 X8      1,2,4,5      .465583E-03
2 KONSUMSI      0.500      .244296E-03
3 LUASKAVL      116.500      .243263E-03
4 X11      1,4,5      .207970E-03
5 LUASLANT      23.500      .197875E-03

```

```

*****
*          Node 38: X11          *
*          N: 1346          *
*****

=====
=          Terminal Node 36      =          Terminal Node 37      =
=          N: 650          =          N: 696          =
=====

```

```

Node 38 was split on X11
A case goes left if X11 = (1,5)
Improvement = .396742E-03      Complexity Threshold = .414970E-03

Node      Cases      Wgt Counts      Cost      Class
 38      1346      1346.00      0.623 1
-36      650      650.00      0.568 1
-37      696      696.00      0.653 2

      Weighted Counts
Class      Top      Left      Right
1          261.00      151.00      110.00
2          536.00      247.00      289.00
3          549.00      252.00      297.00

      Within Node Probabilities
Class      Top      Left      Right
1          0.377      0.432      0.320
2          0.318      0.291      0.347
3          0.305      0.277      0.333

      Surrogate      Split      Assoc.      Improve.
1 X8      s 3,4,5      0.174      .334704E-03
2 X16     s 1,5,6      0.165      .123310E-04

```

```

3 X6          s 3,4,6          0.137 .712742E-05
4 KONSUMSI    s      0.500      0.134 .101770E-03
5 X5          s 2,3          0.109 .207182E-03
Competitor      Split
1 X8          1,2,5          .356115E-03
2 X5          2,3          .207182E-03
3 X10         3,4,6          .196909E-03
4 LUASLANT    29.000         .158316E-03
5 KONSUMSI    1.500         .127394E-03

```

```

*****
*                Node 39: X10                *
*                N: 10826                    *
*****
=====
= Terminal Node 39 = * Node 40 *
= N: 716           = * N: 10110 *
=====
*****

```

Node 39 was split on X10

A case goes left if X10 = (5,6)

Improvement = 0.001 Complexity Threshold = .561584E-03

Node	Cases	Wgt Counts	Cost	Class
39	10826	10826.00	0.590	1
-39	716	716.00	0.477	1
40	10110	10110.00	0.599	1

Weighted Counts

Class	Top	Left	Right
1	2355.00	219.00	2136.00
2	4827.00	319.00	4508.00
3	3644.00	178.00	3466.00

Within Node Probabilities

Class	Top	Left	Right
1	0.410	0.523	0.401
2	0.346	0.314	0.349
3	0.244	0.164	0.250

```

Competitor      Split
1 KEPEMILI      905000.000      Improve.
2 X6            4,6          .920650E-03
3 X5            2,3          .814041E-03
4 X8            4,5          .776999E-03
5 X9            2,3,4        .759588E-03
                     .723636E-03

```

```

*****
*                Node 40: KEPEMILI            *
*                N: 10110                    *
*****
=====
= Terminal Node 40 = * Node 41 *
= N: 2163          = * N: 7947 *
=====
*****

```

Node 40 was split on KEPEMILI

A case goes left if KEPEMILI <= 207500.000

Improvement = 0.001 Complexity Threshold = .673900E-03

Node	Cases	Wgt Counts	Cost	Class
40	10110	10110.00	0.599	1
-40	2163	2163.00	0.539	1
41	7947	7947.00	0.616	1

Weighted Counts

Class	Top	Left	Right
1	2136.00	552.00	1584.00
2	4508.00	926.00	3582.00
3	3466.00	685.00	2781.00

Within Node Probabilities

Class	Top	Left	Right
-------	-----	------	-------

```

1          0.401          0.461          0.384
2          0.349          0.319          0.357
3          0.250          0.220          0.259

Competitor      Split
1 X8            4,5
2 KONSUMSI      0.500
3 X9            2,3,4
4 X6            4,6
5 X7            1,3,4
Improve.
.677305E-03
.638523E-03
.623317E-03
.542893E-03
.520800E-03

*****
*                Node 41: X8                *
*                N: 7947                    *
*****

=====
=      Terminal Node 41      =      *      Node 42      *
=      N: 2752              =      *      N: 5195      *
=====
*****

Node 41 was split on X8
A case goes left if X8 = (2,4,5)
Improvement = .563786E-03      Complexity Threshold = .842379E-03

Node      Cases      Wgt Counts      Cost Class
41         7947         7947.00         0.616 1
-41        2752         2752.00         0.580 1
42         5195         5195.00         0.637 1

Weighted Counts
Class      Top      Left      Right
1          1584.00      619.00      965.00
2          3582.00     1244.00     2338.00
3          2781.00      889.00     1892.00

Within Node Probabilities
Class      Top      Left      Right
1          0.384         0.420      0.363
2          0.357         0.348      0.362
3          0.259         0.232      0.274

Competitor      Split
1 X7            1,3,4
2 KEPEMILI      975000.000
3 X9            2,3,4
4 KONSUMSI      0.500
5 X5            2,3
Improve.
.468097E-03
.416695E-03
.408015E-03
.399861E-03
.360204E-03

*****
*                Node 42: X9                *
*                N: 5195                    *
*****

*****
*                Node 43                *      *      Node 45      *
*                N: 2288                *      *      N: 2907      *
*****

Node 42 was split on X9
A case goes left if X9 = (2,3,4)
Improvement = .526107E-03      Complexity Threshold = 0.001

Node      Cases      Wgt Counts      Cost Class
42         5195         5195.00         0.637 1
43         2288         2288.00         0.601 1
45         2907         2907.00         0.633 2

Weighted Counts
Class      Top      Left      Right
1          965.00         481.00      484.00
2          2338.00      1046.00     1292.00
3          1892.00         761.00     1131.00

```



```

-42      413      413.00      0.599 2
-43      1254     1254.00      0.598 1
      Weighted Counts
Class      Top      Left      Right
1          320.00     54.00     266.00
2          751.00     192.00     559.00
3          596.00     167.00     429.00
      Within Node Probabilities
Class      Top      Left      Right
1          0.373     0.274     0.402
2          0.360     0.401     0.348
3          0.267     0.326     0.250
      Competitor      Split      Improve.
1 PENGHASI      395000.000      .276914E-03
2 X11      1,4,5      .247947E-03
3 X10      1,3,4      .185208E-03
4 V21_A      1.500      .165444E-03
5 KONSUMSI      4.500      .145586E-03

*****
*          Node 45: LUASKAVL          *
*          N: 2907                    *
*****
*****
*          Node 46          *          *          Node 51          *
*          N: 1900          *          *          N: 1007          *
*****
Node 45 was split on LUASKAVL
A case goes left if LUASKAVL <= 134.500
Improvement = .202513E-03      Complexity Threshold = .469094E-03
      Node      Cases      Wgt Counts      Cost      Class
45      2907      2907.00      0.633 2
46      1900      1900.00      0.641 2
51      1007      1007.00      0.618 2
      Weighted Counts
Class      Top      Left      Right
1          484.00     340.00     144.00
2          1292.00     839.00     453.00
3          1131.00     721.00     410.00
      Within Node Probabilities
Class      Top      Left      Right
1          0.334     0.353     0.295
2          0.367     0.359     0.382
3          0.300     0.288     0.323
      Surrogate      Split      Assoc.      Improve.
1 LUASLANT s      97.500      0.073      .560960E-04
      Competitor      Split      Improve.
1 X10      1,2,4      .178926E-03
2 KONSUMSI      0.500      .172611E-03
3 X8      3      .143474E-03
4 X5      2,3      .136669E-03
5 X4      1,2,4,5,6      .131296E-03

*****
*          Node 46: KEPEMILI          *
*          N: 1900                    *
*****
=====
=          Terminal Node 45          =          *          Node 47          *
=          N: 349                    =          *          N: 1551          *
=====
Node 46 was split on KEPEMILI

```

A case goes left if KEPEMILI <= 340000.000

Improvement = .315914E-03 Complexity Threshold = .842154E-03

Node	Cases	Wgt Counts	Cost	Class
46	1900	1900.00	0.641	2
-45	349	349.00	0.596	2
47	1551	1551.00	0.628	1

Weighted Counts

Class	Top	Left	Right
1	340.00	43.00	297.00
2	839.00	162.00	677.00
3	721.00	144.00	577.00

Within Node Probabilities

Class	Top	Left	Right
1	0.353	0.260	0.372
2	0.359	0.404	0.349
3	0.288	0.336	0.278

Competitor	Split	Improve.
1 X10	3	.270589E-03
2 KONSUMSI	0.500	.217678E-03
3 LUASLANT	204.500	.134399E-03
4 LUASKAVL	100.500	.128553E-03
5 X5	2,3	.119308E-03

* Node 47: V21_A *
* N: 1551 *

```

=====
= Terminal Node 46 = * Node 48 *
= N: 67 = * N: 1484 *
=====
*****

```

Node 47 was split on V21_A

A case goes left if V21_A <= 2.500

Improvement = .198298E-03 Complexity Threshold = .412468E-03

Node	Cases	Wgt Counts	Cost	Class
47	1551	1551.00	0.628	1
-46	67	67.00	0.467	1
48	1484	1484.00	0.636	1

Weighted Counts

Class	Top	Left	Right
1	297.00	21.00	276.00
2	677.00	25.00	652.00
3	577.00	21.00	556.00

Within Node Probabilities

Class	Top	Left	Right
1	0.372	0.533	0.364
2	0.349	0.261	0.354
3	0.278	0.205	0.282

Competitor	Split	Improve.
1 KONSUMSI	0.500	.181840E-03
2 X10	3	.178871E-03
3 LUASKAVL	101.500	.170686E-03
4 KEPEMILI	1875000.000	.148054E-03
5 X11	2,3,5	.143085E-03

* Node 48: X10 *
* N: 1484 *

```

=====
= Terminal Node 47 = * Node 49 *
= N: 340 = * N: 1144 *

```

```

===== *****
Node 48 was split on X10
A case goes left if X10 = (3)
Improvement = .223950E-03 Complexity Threshold = .549957E-03

Node      Cases      Wgt Counts      Cost Class
 48      1484      1484.00      0.636 1
-47      340      340.00      0.565 1
 49      1144      1144.00      0.640 2

Weighted Counts
Class      Top      Left      Right
1          276.00      80.00      196.00
2          652.00      150.00      502.00
3          556.00      110.00      446.00

Within Node Probabilities
Class      Top      Left      Right
1          0.364      0.435      0.341
2          0.354      0.336      0.360
3          0.282      0.230      0.299

Competitor      Split      Improve.
1 LUASKAVL      101.500      .184289E-03
2 KONSUMSI      0.500      .163979E-03
3 X11      1,2,3,5      .133054E-03
4 LUASLANT      204.500      .127965E-03
5 X1      1,3,4,5,6      .116065E-03

```

```

*****
*          Node 49: KONSUMSI          *
*          N: 1144                    *
*****
===== *****
=          Terminal Node 48          =          Node 50          *
=          N: 286                    =          N: 858          *
===== *****

```

```

Node 49 was split on KONSUMSI
A case goes left if KONSUMSI <= 0.500
Improvement = .259629E-03 Complexity Threshold = .578210E-03

Node      Cases      Wgt Counts      Cost Class
 49      1144      1144.00      0.640 2
-48      286      286.00      0.614 1
 50      858      858.00      0.611 2

```

```

Weighted Counts
Class      Top      Left      Right
1          196.00      57.00      139.00
2          502.00      99.00      403.00
3          446.00      130.00      316.00

```

```

Within Node Probabilities
Class      Top      Left      Right
1          0.341      0.386      0.326
2          0.360      0.276      0.389
3          0.299      0.338      0.285

```

```

Surrogate      Split      Assoc.      Improve.
1 V20_A s      0.500      0.060      .636558E-04
2 X10 s 2      0.016      .371240E-04

Competitor      Split      Improve.
1 KEPEMILI      1550000.000      .225269E-03
2 LUASKAVL      129.000      .172723E-03
3 LUASLANT      76.500      .132434E-03
4 X4      3,4,6      .124905E-03
5 X11      1,2,3,5      .118194E-03

```

```

*****
*          Node 50: LUASLANT          *
*          N: 858                    *

```



```

*****
=====
=      Terminal Node 49      =      Terminal Node 50      =
=      N: 774                =      N: 84                =
=====

```

Node 50 was split on LUASLANT

A case goes left if LUASLANT <= 76.500

Improvement = .203621E-03 Complexity Threshold = .407175E-03

Node	Cases	Wgt	Counts	Cost	Class
50	858		858.00	0.611	2
-49	774		774.00	0.598	2
-50	84		84.00	0.529	1

Weighted Counts

Class	Top	Left	Right
1	139.00	117.00	22.00
2	403.00	371.00	32.00
3	316.00	286.00	30.00

Within Node Probabilities

Class	Top	Left	Right
1	0.326	0.308	0.471
2	0.389	0.402	0.282
3	0.285	0.290	0.247

Competitor	Split	Improve.
1 LUASKAVL	101.500	.185310E-03
2 KEPEMILI	.107500E+08	.159657E-03
3 X7	1,3,4	.133261E-03
4 X10	2,4	.117547E-03
5 X1	1,4,5,6	.106619E-03

```

*****
*      Node 51: LUASKAVL      *
*      N: 1007                *
*****

```

```

=====
=      Terminal Node 51      =      *      Node 52      *
=      N: 594                =      *      N: 413      *
=====

```

Node 51 was split on LUASKAVL

A case goes left if LUASKAVL <= 232.500

Improvement = .372192E-03 Complexity Threshold = .517113E-03

Node	Cases	Wgt	Counts	Cost	Class
51	1007		1007.00	0.618	2
-51	594		594.00	0.594	2
52	413		413.00	0.632	1

Weighted Counts

Class	Top	Left	Right
1	144.00	66.00	78.00
2	453.00	273.00	180.00
3	410.00	255.00	155.00

Within Node Probabilities

Class	Top	Left	Right
1	0.295	0.239	0.368
2	0.382	0.406	0.350
3	0.323	0.355	0.282

Surrogate	Split	Assoc.	Improve.
1 X5	s 1,3	0.051	.254896E-03
2 LUASLANT	s 205.000	0.025	.303619E-04
3 PENGHASI	s 1075000.000	0.017	.693960E-04

4	KEPEMILI	s	.106175E+08		0.014	.403795E-04
5	KONSUMSI	s	2.500		0.013	.869554E-04
	Competitor		Split			Improve.
1	KEPEMILI		905000.000			.256025E-03
2	X5		1,3			.254896E-03
3	V20_A		1.500			.237405E-03
4	KONSUMSI		1.500			.158040E-03
5	X16		1,3,5,6			.126002E-03

```

*****
*           Node 52: KEPEMILI           *
*           N: 413                       *
*****

```

```

=====
=      Terminal Node 52      =      Terminal Node 53      =
=           N: 216          =      N: 197          =
=====

```

Node 52 was split on KEPEMILI

A case goes left if KEPEMILI <= 950000.000

Improvement = .567245E-03 Complexity Threshold = .854889E-03

Node	Cases	Wgt	Counts	Cost	Class
52	413		413.00	0.632	1
-52	216		216.00	0.527	1
-53	197		197.00	0.567	2

Weighted Counts

Class	Top	Left	Right
1	78.00	57.00	21.00
2	180.00	84.00	96.00
3	155.00	75.00	80.00

Within Node Probabilities

Class	Top	Left	Right
1	0.368	0.473	0.230
2	0.350	0.287	0.433
3	0.282	0.240	0.337

Surrogate	Split	Assoc.	Improve.
1 PENGHASI	s 775000.000	0.179	.104943E-04
2 X6	s 1,2,3,5	0.125	.292791E-04
3 X5	s 1,3	0.123	.184064E-03
4 X10	s 1,3,4	0.078	.898931E-04
5 LUASLANT	s 73.000	0.067	.385749E-04
Competitor	Split		Improve.
1 LUASKAVL	235.500		.240985E-03
2 X5	1,3		.184064E-03
3 X16	4,5,6		.167712E-03
4 X7	1,4		.165939E-03
5 V20_A	1.500		.135509E-03

Output C3. Informasi Proporsi Setiap Kelas pada Setiap Simpul Pohon Klasifikasi Optimal

=====

TERMINAL NODE INFORMATION

=====

[Breiman adjusted cost, lambda = .337166E-03]

Node	Class	Wgt Count	N	Prob	Cost	Parent Complexity
1	3	567.00	567	0.012	0.567	.427246E-03 [0.614]
1		68.00	68	0.256		
2		200.00	200	0.310		
3		299.00	299	0.433		
2	2	784.00	784	0.018	0.651	.427246E-03 [0.684]
1		127.00	127	0.327		
2		329.00	329	0.349		
3		328.00	328	0.325		
3	1	385.00	385	0.009	0.585	0.002 [0.647]
1		85.00	85	0.415		
2		166.00	166	0.334		
3		134.00	134	0.252		
4	3	4574.00	4574	0.100	0.584	.771530E-03 [0.590]
1		593.00	593	0.273		
2		1635.00	1635	0.310		
3		2346.00	2346	0.416		
5	3	7300.00	7300	0.155	0.558	.514349E-03 [0.562]
1		786.00	786	0.234		
2		2649.00	2649	0.324		
3		3865.00	3865	0.442		
6	3	475.00	475	0.010	0.494	0.002 [0.553]
1		42.00	42	0.197		
2		153.00	153	0.296		
3		280.00	280	0.506		
7	1	82.00	82	0.002	0.449	.675886E-03 [0.681]
1		27.00	27	0.551		
2		30.00	30	0.252		
3		25.00	25	0.196		
8	2	789.00	789	0.017	0.619	.576914E-03 [0.653]
1		102.00	102	0.271		
2		348.00	348	0.381		
3		339.00	339	0.347		
9	3	147.00	147	0.003	0.488	.576914E-03 [0.670]
1		11.00	11	0.170		
2		50.00	50	0.318		
3		86.00	86	0.512		
10	1	47.00	47	0.001	0.485	.440466E-03 [0.864]
1		14.00	14	0.515		
2		18.00	18	0.273		

3	15.00	15	0.212		
11 2	341.00	341	0.008	0.624	.681117E-03 [0.700]
1	47.00	47	0.286		
2	150.00	150	0.376		
3	144.00	144	0.337		
12 1	1385.00	1385	0.034	0.576	.681117E-03 [0.593]
1	315.00	315	0.424		
2	588.00	588	0.326		
3	482.00	482	0.250		
13 1	88.00	88	0.003	0.311	.515778E-03 [0.506]
1	41.00	41	0.689		
2	15.00	15	0.104		
3	32.00	32	0.207		
14 3	325.00	325	0.007	0.615	.515778E-03 [0.694]
1	50.00	50	0.315		
2	116.00	116	0.301		
3	159.00	159	0.385		
15 1	21.00	21	.642023E-03	0.353	.494289E-03 [0.969]
1	9.00	9	0.647		
2	11.00	11	0.326		
3	1.00	1	0.028		
16 2	209.00	209	0.005	0.571	.591051E-03 [0.693]
1	27.00	27	0.270		
2	104.00	104	0.429		
3	78.00	78	0.301		
17 3	643.00	643	0.015	0.632	.591051E-03 [0.672]
1	102.00	102	0.322		
2	238.00	238	0.310		
3	303.00	303	0.368		
18 1	258.00	258	0.006	0.551	.467661E-03 [0.639]
1	63.00	63	0.449		
2	83.00	83	0.244		
3	112.00	112	0.307		
19 2	669.00	669	0.015	0.604	.823470E-03 [0.644]
1	84.00	84	0.265		
2	305.00	305	0.396		
3	280.00	280	0.340		
20 3	2895.00	2895	0.064	0.612	.823470E-03 [0.622]
1	410.00	410	0.294		
2	1079.00	1079	0.318		
3	1406.00	1406	0.388		
21 2	1961.00	1961	0.044	0.607	0.005 [0.621]
1	272.00	272	0.287		
2	903.00	903	0.393		
3	786.00	786	0.320		
22 1	177.00	177	0.007	0.140	.519674E-03 [0.225]
1	126.00	126	0.860		

	2	35.00	35	0.098		
	3	16.00	16	0.042		
23	2	101.00	101	0.002	0.470	.496640E-03 [0.696]
	1	16.00	16	0.317		
	2	65.00	65	0.530		
	3	20.00	20	0.153		
24	1	95.00	95	0.003	0.301	.496640E-03 [0.479]
	1	46.00	46	0.699		
	2	34.00	34	0.213		
	3	15.00	15	0.088		
25	2	427.00	427	0.010	0.604	.979957E-03 [0.665]
	1	59.00	59	0.286		
	2	198.00	198	0.396		
	3	170.00	170	0.318		
26	3	1291.00	1291	0.029	0.632	.510987E-03 [0.653]
	1	175.00	175	0.283		
	2	524.00	524	0.349		
	3	592.00	592	0.368		
27	1	98.00	98	0.003	0.513	.466898E-03 [0.721]
	1	27.00	27	0.487		
	2	41.00	41	0.305		
	3	30.00	30	0.208		
28	2	82.00	82	0.002	0.470	.466898E-03 [0.768]
	1	6.00	6	0.164		
	2	47.00	47	0.530		
	3	29.00	29	0.306		
29	1	688.00	688	0.016	0.626	.553526E-03 [0.662]
	1	132.00	132	0.374		
	2	245.00	245	0.286		
	3	311.00	311	0.339		
30	2	520.00	520	0.011	0.578	.413651E-03 [0.632]
	1	48.00	48	0.203		
	2	242.00	242	0.422		
	3	230.00	230	0.375		
31	3	46.00	46	.885912E-03	0.319	.413651E-03 [0.811]
	1	2.00	2	0.104		
	2	10.00	10	0.214		
	3	34.00	34	0.681		
32	2	143.00	143	0.003	0.439	.618175E-03 [0.631]
	1	6.00	6	0.098		
	2	83.00	83	0.561		
	3	54.00	54	0.341		
33	3	367.00	367	0.007	0.516	.618175E-03 [0.594]
	1	27.00	27	0.167		
	2	137.00	137	0.349		
	3	203.00	203	0.484		
34	1	28.00	28	.761323E-03	0.455	.410204E-03

					[1.003]
	1	9.00	9	0.545	
	2	7.00	7	0.175	
	3	12.00	12	0.280	
35	3	1023.00	1023	0.023	0.599 0.002 [0.625]
	1	156.00	156	0.313	
	2	347.00	347	0.286	
	3	520.00	520	0.401	
36	1	650.00	650	0.016	0.568 .414970E-03 [0.605]
	1	151.00	151	0.432	
	2	247.00	247	0.291	
	3	252.00	252	0.277	
37	2	696.00	696	0.016	0.653 .414970E-03 [0.691]
	1	110.00	110	0.320	
	2	289.00	289	0.347	
	3	297.00	297	0.333	
38	1	2341.00	2341	0.057	0.596 0.001 [0.607]
	1	498.00	498	0.404	
	2	997.00	997	0.333	
	3	846.00	846	0.264	
39	1	716.00	716	0.019	0.477 .561584E-03 [0.508]
	1	219.00	219	0.523	
	2	319.00	319	0.314	
	3	178.00	178	0.164	
40	1	2163.00	2163	0.055	0.539 .673900E-03 [0.550]
	1	552.00	552	0.461	
	2	926.00	926	0.319	
	3	685.00	685	0.220	
41	1	2752.00	2752	0.068	0.580 .842379E-03 [0.589]
	1	619.00	619	0.420	
	2	1244.00	1244	0.348	
	3	889.00	889	0.232	
42	2	413.00	413	0.009	0.599 0.001 [0.663]
	1	54.00	54	0.274	
	2	192.00	192	0.401	
	3	167.00	167	0.326	
43	1	1254.00	1254	0.031	0.598 0.001 [0.617]
	1	266.00	266	0.402	
	2	559.00	559	0.348	
	3	429.00	429	0.250	
44	1	621.00	621	0.016	0.535 .578034E-03 [0.572]
	1	161.00	161	0.465	
	2	295.00	295	0.351	
	3	165.00	165	0.184	
45	2	349.00	349	0.008	0.596 .842154E-03 [0.672]
	1	43.00	43	0.260	
	2	162.00	162	0.404	
	3	144.00	144	0.336	

46	1	67.00	67	0.002	0.467	.412468E-03 [0.746]
	1	21.00	21	0.533		
	2	25.00	25	0.261		
	3	21.00	21	0.205		
47	1	340.00	340	0.008	0.565	.549957E-03 [0.634]
	1	80.00	80	0.435		
	2	150.00	150	0.336		
	3	110.00	110	0.230		
48	1	286.00	286	0.007	0.614	.578210E-03 [0.699]
	1	57.00	57	0.386		
	2	99.00	99	0.276		
	3	130.00	130	0.338		
49	2	774.00	774	0.018	0.598	.407175E-03 [0.632]
	1	117.00	117	0.308		
	2	371.00	371	0.402		
	3	286.00	286	0.290		
50	1	84.00	84	0.002	0.529	.407175E-03 [0.771]
	1	22.00	22	0.471		
	2	32.00	32	0.282		
	3	30.00	30	0.247		
51	2	594.00	594	0.013	0.594	.517113E-03 [0.640]
	1	66.00	66	0.239		
	2	273.00	273	0.406		
	3	255.00	255	0.355		
52	1	216.00	216	0.006	0.527	.854889E-03 [0.629]
	1	57.00	57	0.473		
	2	84.00	84	0.287		
	3	75.00	75	0.240		
53	2	197.00	197	0.004	0.567	.854889E-03 [0.700]
	1	21.00	21	0.230		
	2	96.00	96	0.433		
	3	80.00	80	0.337		